PL-TR-91-2098

AD-A239 179

(2)

CFLOS4D ACCURACY ASSESSMENT USING WHOLE SKY IMAGER DATA

KENNETH B. MACNICHOL
STEVEN R. FINCH

TASC
55 Walkers Brook Drive
Reading, MA 01867

18 April 1991

Final Report
1 September 1988 — 30 April 1991

DTIC
ELECTE
AUG 07 1991
S B D

Approved for public release; distribution unlimited

PHILLIPS LABORATORY
AIR FORCE SYSTEMS COMMAND
HANSCOM AIR FORCE BASE, MASSACHUSETTS 01731-5000

91-06933

This technical report has been reviewed and is approved for publication.

BRUCE A. KUNKEL
Contract Manager

DONALD D. GRANTHAM
Branch Chief

ROBERT A. McCLATCHEY
Division Director

This document has been reviewed by the ESD Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS).

Qualified requestors may obtain additional copies from the Defense Technical Information Center. All others should apply to the National Technical Information Service.

If your address has changed, or if you wish to be removed from the mailing list, or if the addressee is no longer employed by your organization, please notify PL/IMA, Hanscom AFB, MA 01731-5000. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document requires that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	18 April 1991	Final 9/1/88 to 4/30/91

4. TITLE AND SUBTITLE

CFLOS4D Accuracy Assessment Using Whole Sky Imager Data

5. FUNDING NUMBERS

PE 62101F

P6670 T09 WUAR

Contract F19628-88-C-0153

6. AUTHOR(S)

Kenneth B. MacNichol
Steven R. Finch

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

TASC
55 Walkers Brook Drive
Reading, MA 01867

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Phillips Laboratory
Hanscom AFB, MA 01731-5000

Contract Manager: Bruce Kunkel/LYA

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

PL-TR-91-2098

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The CFLOS4D and CFARC simulation programs are based on several fundamental models of clouds and cloud cover effects. Validating these programs, therefore, requires an approach which recognizes their inherent hierarchical nature. This report describes the hierarchical model validation and associated analyses of the CFLOS4D program using data collected by the WSI system. Specifically, we discuss the rigorous methodology used to compare WSI data-derived results with simulator predictions at the system level and to perform data requirements analyses which determine the quantity of WSI data needed to assess simulator utility within specified accuracy. We then discuss several investigations at the model component level, including models governing the probability of a cloud-free line-of-sight (CFLOS), CFLOS recurrence and persistence frequencies and sky cover temporal and spatial correlations.

14. SUBJECT TERMS

cloud-free line-of-sight
data requirements analysis
model validation
Monte Carlo simulation

15. NUMBER OF PAGES
76

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

FOREWORD

Under contract to the Geophysics Laboratory, TASC has undertaken an effort to validate the CFLOS4D (Cloud-Free Line-of-Sight, 4 Dimensions) and CFARC (Cloud-Free Arcs) simulation programs. These simulators provide realizations of downtime durations (due to cloud obscuration) of ground-based laser systems. The validation activity is based on a 'truth' cloud image data set provided by the Marine Physical Laboratory of the Scripps Institute of Oceanography using a new instrument — the Whole Sky Imager (WSI).

Since funding for both WSI data collection and the final portion of the project was cut due to changing government priorities, the original validation plan could not be carried out completely. Instead, concluding activity for this project has been directed toward

- completing validation of the CFLOS4D simulator using a limited, preliminary WSI data set
- determining the bounds of statistical confidence and the power in the validation procedure applied to existing data
- assessing the improvement in these bounds, were a larger WSI data set to become available.

Considerable effort has been devoted to properly interpreting the validation results, with the understanding that recommendations in this report should be instrumental in guiding further WSI data collection and analysis.

Accession For
NTIS GRA&I ☑
DTIC TAB ☐
Unannounced ☐
Justification
By
Distribution/
Availability Codes
Dist | Avail and/or Special
A-1

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF FIGURES (Continued)

LIST OF FIGURES (Continued)

LIST OF TABLES

1. INTRODUCTION

This section provides background information and motivation for the validation effort. The CFLOS simulators and their outputs are briefly described, as is the WSI data set. Project objectives as well as the purpose and scope of this report are discussed in the concluding subsection.

1.1 BACKGROUND

By providing model-produced realizations of downtime* durations, the CFLOS simulators provide a means to study and assess the ability of ground-based laser (GBL) systems to maintain line-of-sight to a space-based relay mirror. In particular, effectiveness studies involving the locations and number of potential GBL sites can be readily executed. For example, a critical question addressed by the simulators is the number of GBL sites required to reduce downtime to a prescribed acceptable level. Simulation output is organized to readily answer such questions.

CFLOS4D simulates cloud-free line-of-sight to a geostationary satellite. Downtime is due only to cloud obscuration. CFARC simulates cloud-free arcs to orbiting satellites, therefore downtime includes those times for which the path to a satellite is obstructed by the earth. Both programs generate downtime durations for one site or for systems of several sites.

In addition to site and satellite orbit parameters, the major program inputs are mean sky cover and scale distance data which define the climatological sky cover distribution for each site. Mean sky cover and scale distance are computed from other cloud models developed by the Geophysics Laboratory (Ref. 1); previously prepared site-specific data files based on those models have been provided to TASC for the validation effort.

The basic output of each program is a table, functionalized by downtime duration categories (e.g., 1-5 minutes down is one category, 6-30 minutes is another, etc.) and site configuration (e.g., a one-site system, a two-site system, etc.), indicating the number of times each system was down within each category. The percentage of total simulated time that a system is down in each category is also provided. Usually, the simulated time span for a run is an integral number of years, with time steps of one minute. A good description of program inputs, outputs, and operations is provided in the CFLOS User's Manual (Ref. 2).

*The simulators permit two definitions of system 'down'. Most commonly, a system of several sites is down when all sites are down (i.e., all sites have lost line-of-sight). Only this definition is employed for this validation effort.

The various statistical model components underlying CFLOS simulations are described in Ref. 3. These 'subsystem' components, such as site-to-site spatial correlation models, CFLOS temporal correlation models and the single point CFLOS probability model are in themselves subjects of the validation effort. Relationships between subsystem components are illustrated in Fig. 1.1-1 and are described further in Section 4. Thus this report divides naturally into two parts: a discussion of 'system level' investigations (involving the tables of downtime duration statistics for various time interval categories and for various ground site locations) and a discussion of 'model component' investigations (involving the subsystem models mentioned above).

To support validation of the CFLOS models, digital cloud image data, collected from the new Whole Sky Imager, have been acquired from the Marine Physical Laboratory of the Scripps Institute under the sponsorship of the Geophysics Laboratory. The Whole Sky Imager is an automated ground-based electronic imaging system which gathers calibrated multi-spectral images of the sky dome (Ref. 4). Seven Whole Sky Imagers were in operation across the U.S. as of January 1990, collecting data for cloud/no-cloud processing. The seven WSI sites were at White Sands C-Station, New Mexico; Kirtland AFB, New Mexico; Columbia, Missouri; Malmstrom AFB, Montana; Malabar AFB, Florida; White Sands HELSTF, New Mexico; and China Lake, California. Very fine sky dome spatial (1/3 degree) and temporal (one minute) resolutions characterize the data as unique.

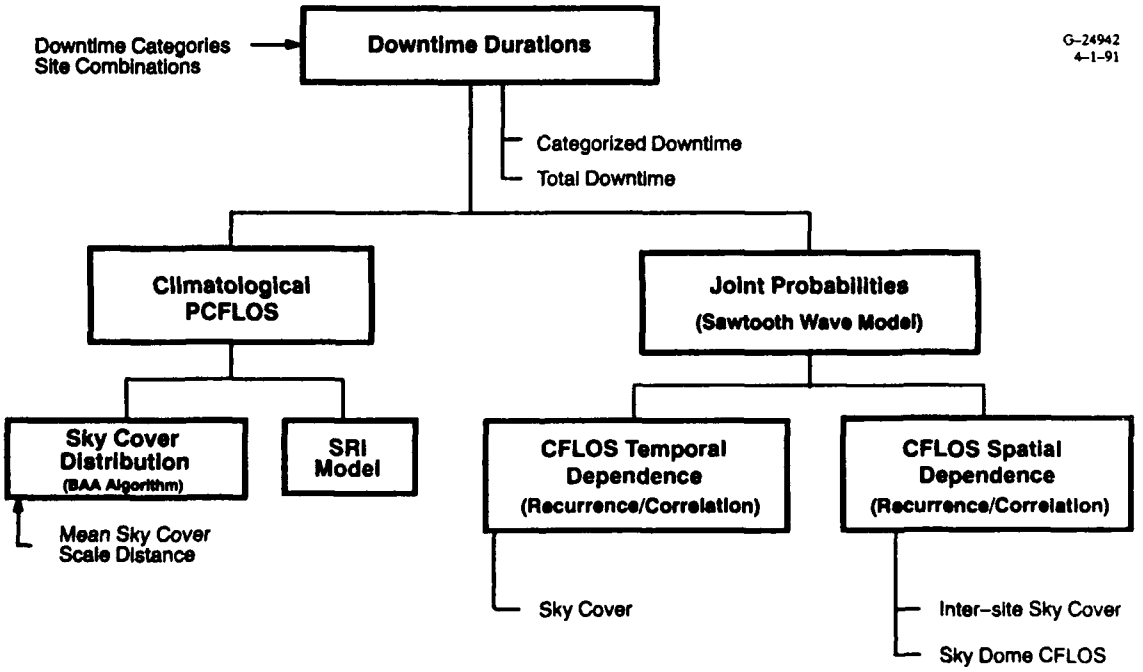


Figure 1.1-1 Hierarchical Validation Components

The nature of the Imager is such that data are collected during daylight hours only (Fig. 1.1-2); moreover, for reasons of data quality, we restrict attention only to the period beginning one hour after sunrise and ending one hour before sunset. Data sequences obtained, therefore, are not continuous. Furthermore, data gaps caused by unknown events are common. Days in which data gaps occur have been eliminated from consideration.

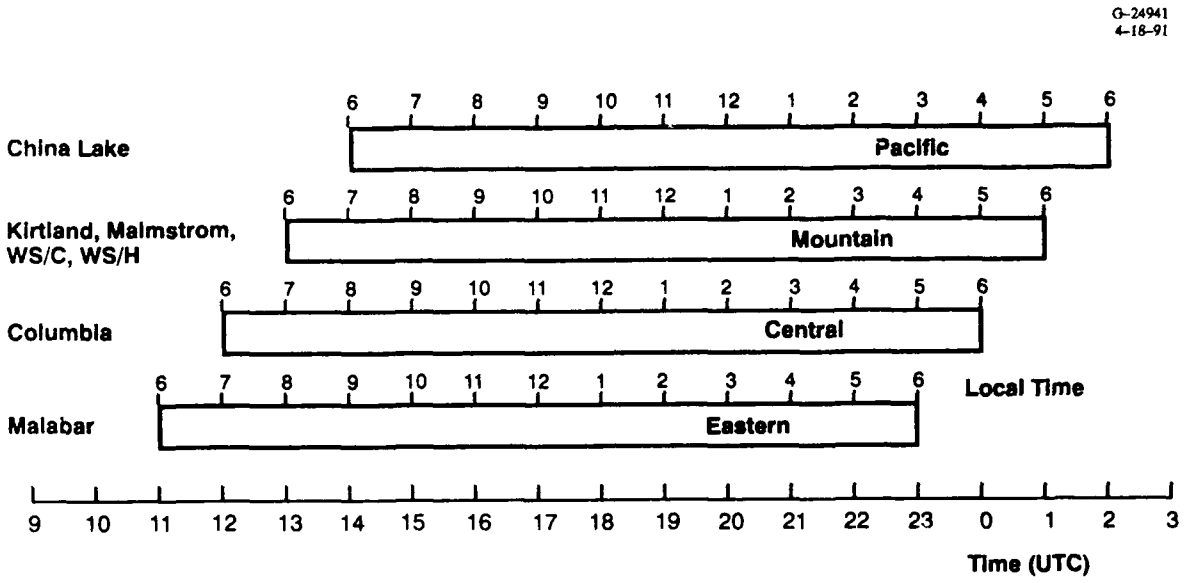


Figure 1.1-2 WSI Data Collection Timelines

1.2 PURPOSE AND SCOPE OF THIS REPORT

The accuracy of the CFLOS simulators must be established if their predictions are to be used with confidence. TASC's directive for this program is to assess the validity of the simulators, using the WSI "truth" data set and assuming it to be error-free. Although directed not to assess WSI data accuracy, TASC recognizes that conclusions regarding model validity could be affected by significant data errors. We have therefore been very conservative so as to select data most likely to be free of processing errors.

System level validation, that is, validation of the model output, is of prime concern (Section 3). However, identification and validation of subsystem models is an important part of our hierarchical validation approach (Section 4). Much of this work pertains to both CFLOS4D and CFARC since the two simulators share many model components (we have not, however, addressed CFARC at the system level). In addition, a data requirements analysis is useful to provide a realistic assessment of the amount of WSI data required to validate (or invalidate) the simulation models at prescribed statistical confidence and power levels (Section 3).

Section 2 contains a chronology of our system level and model component statistical validation investigations, and serves as an overview of the project effort. The system level validation approach, discussed in detail in Section 3, is similar to the data requirements analysis approach presented in Ref. 5 in several respects. Both are rooted in the Monte Carlo simulation of multiple iterations of the discrete Kolmogorov-Smirnov goodness-of-fit test. Both compare the observed WSI histograms (one for each selected downtime duration category) with the corresponding CFLOS4D histograms using various comparison criteria; thus both give rise to not just a single quantity (e.g., a decision whether to accept or reject consistency, or a required sample size) but rather a family of curves. Such curves carry a great deal of information, allowing the analyst to understand the *sensitivity* of decision-making or of the minimum required data set size to a number of parameters. Section 4 describes results of a number of model component activities which provide explanations for certain model/data discrepancies observed at the system level. Section 5 presents conclusions and makes recommendations on future WSI data collection in view of reduced program funding.

2. SUMMARY OF PROJECT ACTIVITY

In this section, a chronology of the events significant to this project is given. One purpose of this discussion is to provide a sense of the progress achieved and difficulties overcome during the contract period. Another purpose of this discussion is to present results which are important and interesting yet somewhat outside the scope of the original contract, specifically

- Comparisons of WSI data and surface observations (in essence a data quality investigation)
- Model impacts on GBL siting (a model application discussion).

2.1 INITIAL INVESTIGATIONS AND DESIGN

In the early months of the project (August through October, 1988) much attention was devoted to the proper formulation of a system level validation approach (since it would dictate the WSI data requirements for accurate model validation). The classical t test and F test for, respectively, the mean and variance of day-to-day downtime duration counts (for a specified downtime duration category) were initially considered. Such tests, however, depend on downtime duration counts following a normal distribution and thus were soon discarded. A test based on the ratio of two Poisson means formed the basis of our preliminary data requirements analysis (Refs. 6 and 7). The attractive features of this approach, which assumed that day-to-day downtime duration counts are approximately Poisson distributed, included:

- Analytic expressions for the minimum required data sample size, functionalized by specified false rejection and false acceptance probabilities and by specified model/data error tolerance ratios
- The ability to perform a reasonable data requirements assessment for a variety of error probabilities and model/data error tolerance ratios, even though WSI data were not available.

Ultimately, results of multiple CFLOS4D runs (presented in Ref. 8) indicated that the downtime counts are not accurately represented by the Poisson distribution assumption and that downtime counts are larger and more variable than had been earlier thought (on the basis of Ref. 2). These results led to the nonparametric, Monte Carlo - based procedure for system level validation which was actually used (Section 2.3).

By February 1989, a model validation database and analysis system was designed. Based on the requirements, size and scope of this project, the decision was made to develop a file-oriented database management system on the project computer, a COMPAQ 80386 PC, rather than use a commercial database (Figs. 2-1 and 2-2). Three main objectives were addressed in the database design: minimize disk

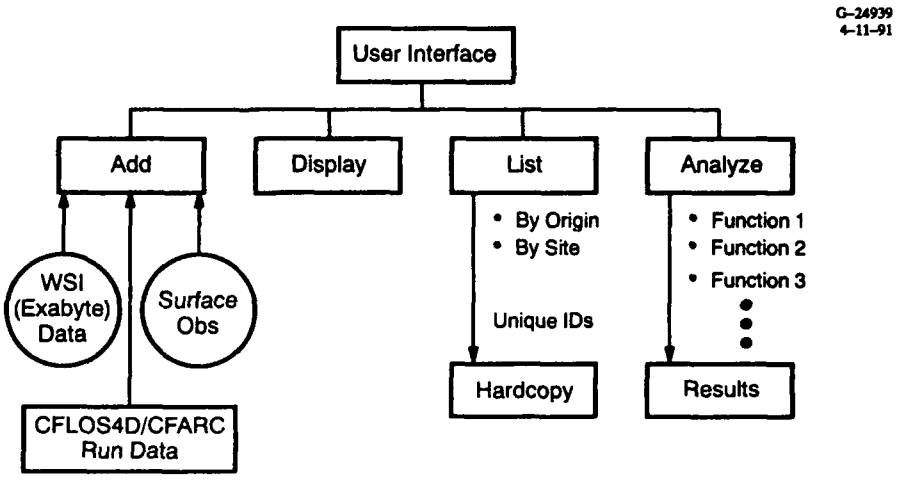


Figure 2-1 CFLOS Database Management System

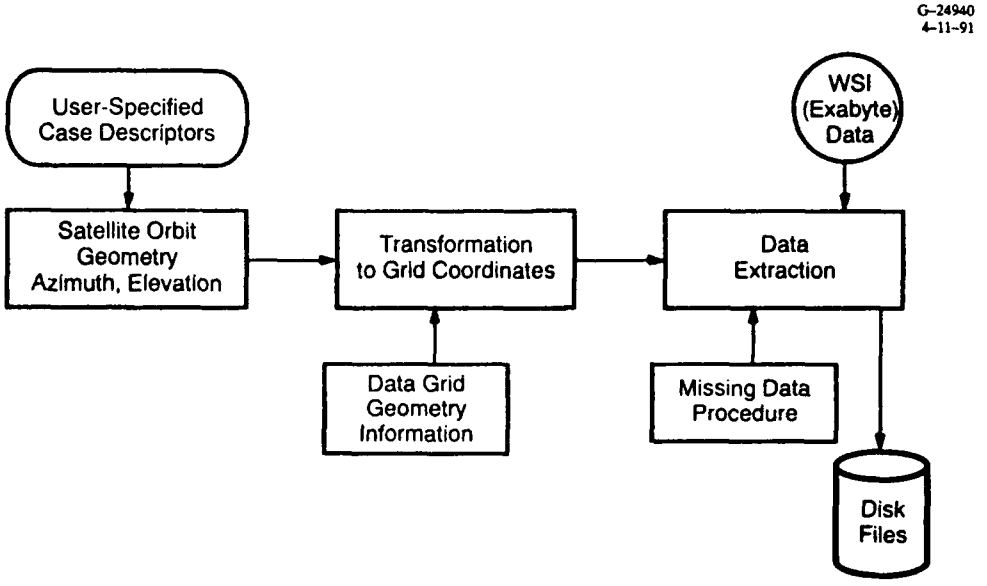


Figure 2-2 WSI Data Extraction Process

space requirements, provide rapid data retrieval, and organize the data so as to permit flexibility in constructing the validation tests.

At the same time, a study of CFLOS4D day-to-day downtime duration count temporal correlation coefficients (computed by the product-moment formula) indicated that day-to-day temporal dependence was insignificant, for any number of sites. This conclusion was confirmed by use of the chi squared test for independence. A simple exponential model for inter-pixel spatial correlation was also developed. The effective increase in WSI sample size (i.e., number of data days), if multiple pixels were utilized per WSI image in the validation procedure, was observed to be significant only if sky dome spatial correlations were sufficiently small. The use of multiple pixels per image is equivalent to the simultaneous consideration of different orbiting satellite locations. For system level validation several sites are involved, and the spatial correlation of time series corresponding to different satellite locations is expected to be smaller than the single site sky dome spatial correlation.

Procedures for comparing simulator model components with corresponding characteristics of WSI data were developed during the period from February through December 1989 as well. These procedures included algorithms for estimating:

- minute-to-minute sky cover temporal correlation
- site-to-site sky cover spatial correlation
- CFLOS persistence and recurrence frequencies
- CFLOS temporal correlation
- cloudy interval distributions
- PCFLOS at designated points in the sky dome.

One issue raised, for example, was whether the two-category limitation (*clear* if sky cover is less than or equal to 50%; *cloudy* otherwise) in the tetrachoric estimate of temporal correlation introduced significant distortion in models underlying CFLOS simulation. A comparison of four-category polychoric correlation estimates with tetrachoric estimates demonstrated a very reasonable fit and thus more than two categories are unnecessary for computing temporal correlation. Another issue concerned a technical error (involving the computation of conditional probabilities) in Ref. 9 which forms the basis for certain CFLOS models of persistence and recurrence. The error was confirmed and solution verified at GL; corrections appear in Ref. 10.

2.2 DETAILED STATISTICAL TEST DESIGN AND INITIAL DATA ANALYSIS

During the period May through July 1989, efforts continued to develop a distribution-free system level validation test, beginning with use of the Wilcoxon-Mann-Whitney (WMW) test for means and culminating with use of the discrete Kolmogorov-Smirnov (KS) test coupled with Monte Carlo routines to compute appropriate test thresholds. Such nonparametric tests have the advantage of generality over parametric tests (e.g., the Poisson-based test) but at the cost of some discrimination power. The WMW test was considered because certain theoretical tools, known as asymptotic relative efficiency ratios, permitted one to transform from a parametric sample size (computed, e.g., as for the Poisson-based test) to a nonparametric sample size. It was determined that roughly 10% more data were required by the WMW procedure than by the Poisson-based procedure to test with the same degree of confidence. The KS test was finally adopted because it is more standard, it is sensitive to a wide variety of distribution deviations, and it is recommended by Refs. 11 and 12.

The first shipment of WSI cloud/no-cloud images, called a 'test sample', arrived at TASC in August 1989. Previously developed software for reading, processing and displaying the images was modified to accommodate a new format. By September, the remaining portion of the test sample WSI data set was received. Much effort was expended in reviewing the images to ascertain the degree of data quality. Cloud detection problems were evident in the vicinity of the solar occultor, near the horizon, and during the time periods near sunrise and sunset. The following quality control guidelines were established, with GL approval, regarding the data extracted for use in our analyses:

- avoid elevation angles less than 20°
- avoid sunrise/sunset time periods (by approximately one hour)
- avoid data points near the occultor.

Data of suspect quality were flagged for comparison with surface observations. Numerous modifications were made to previously developed TASC software to detect and overcome various data problems (e.g., data gaps, data header inconsistencies, etc.).

A TASC memorandum (Ref. 8) describing the KS test procedure for system level validation and consequential data requirements was completed in November 1989. Since only 25 days of the 4 month sample WSI data set were deemed usable for the four WSI sites considered as a system, WSI data results were not presented in detail in that memorandum or in the subsequent CIDOS conference presentation in early January 1990 (Ref. 5).

In February 1990 a meeting involving GL, TASC and GBL Program Office representatives was held at TASC to exchange information on CFLOS simulator use and validation. It was determined that a

broad range of simulator predictions should be validated, including various downtime duration categories and different site configurations. The four downtime duration categories are 1-5 minutes, 6-30 minutes, 31-181 minutes and greater-than-three hours. It was indicated that satellite subpoint locations should correspond to high elevation angles. Current GBL system availability requirements defining allowable outage durations due to cloud obscuration, for specific site configurations were discussed. System outages of three hours or more are of particular concern, as such events could be forecast with reasonable accuracy and impose a window of vulnerability. One outcome of this meeting was a decision to extend the system level validation procedure to include downtime percentage per day for each of the four downtime categories as validation quantities.

Project activity increased during March 1990 in anticipation of the receipt of the first delivery of final quality controlled WSI data (which was to be used for validation). Software which generates histograms of categorized downtime counts and downtime percentages per day (for both CFLOS simulator results and WSI data) and which uses such histograms in the KS validation algorithm was finalized and tested. Software was also developed to determine, for each month and for each system of sites (spanning up to four time zones), the maximum daily time span, in hours, of valid WSI data. It was found, for instance, that systems comprised of sites in the Pacific and Central time zones possessed valid data ranges of 10 hours in July but only 5 hours in January. Such results limit the total data usable for multi-site system validation.

2.3 DATA ANALYSIS

Final quality controlled WSI data were received in monthly shipments from April to September 1990. Much effort was applied to loading the data and revising image processing software to account for changes in header format and content and for revised equations pertaining to image geometry (specifically, the mapping of pixel location to sky dome location). We point out that MPL regards these data as preliminary. MPL plans to employ a direction-dependent thresholding algorithm in the future to improve cloud detection in the occultor and horizon 'problem areas'. (Since the pixels of interest to us in our analyses are in the northern part of the sky dome, away from the horizon and the occultor, our validation results would not be significantly affected by the new procedure.) A run length encoding scheme for compressing the WSI images was additionally implemented to enable efficient transfer of image data between the primary and secondary project computers via floppy disk. The primary machine handled all data loading, image display and system level validation tasks while the secondary machine was used for CFLOS simulator runs and model component validation tasks.

2.3.1 Comparisons with Surface Observations

Project activity at this time included performing comparisons of WSI sky cover data with concurrent surface observations collected at or near a WSI site. The surface observations were in "Airways" format; only clear, scattered, broken and overcast sky cover categories were available from ETAC data files. Figures 2-3 and 2-4 present comparisons of sky cover estimates determined from WSI images with surface observations. Indicated in the figures, which represent data from the Columbia and Kirtland sites, are the mean, mode, and 10/90 percentile bounds for the distribution of WSI-derived sky cover categorized by the corresponding surface observation. WSI images at ten minutes before the hour were used for these comparisons. All daytime data presently available during the period from February through December 1989 were used.

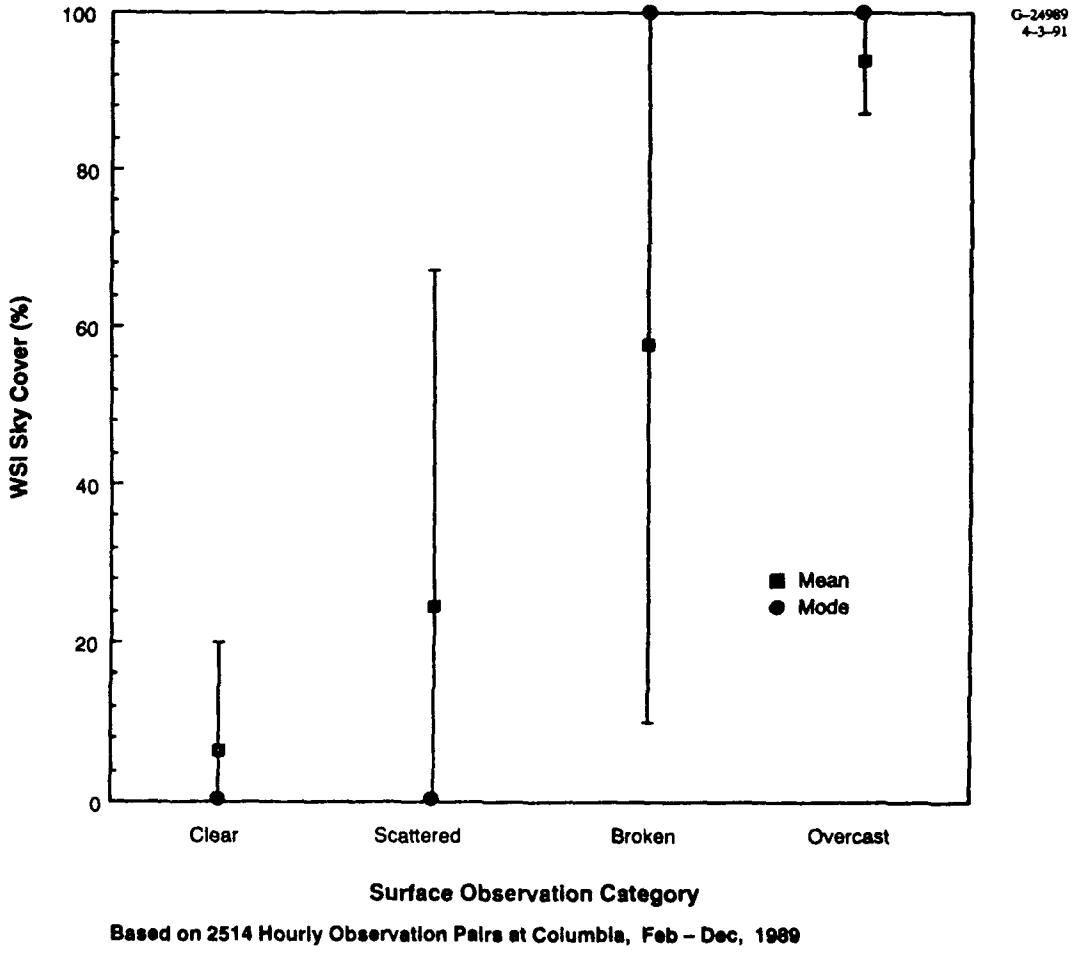


Figure 2-3 WSI/Surface Observation Comparison — Columbia 10/90 Percentile Bounds

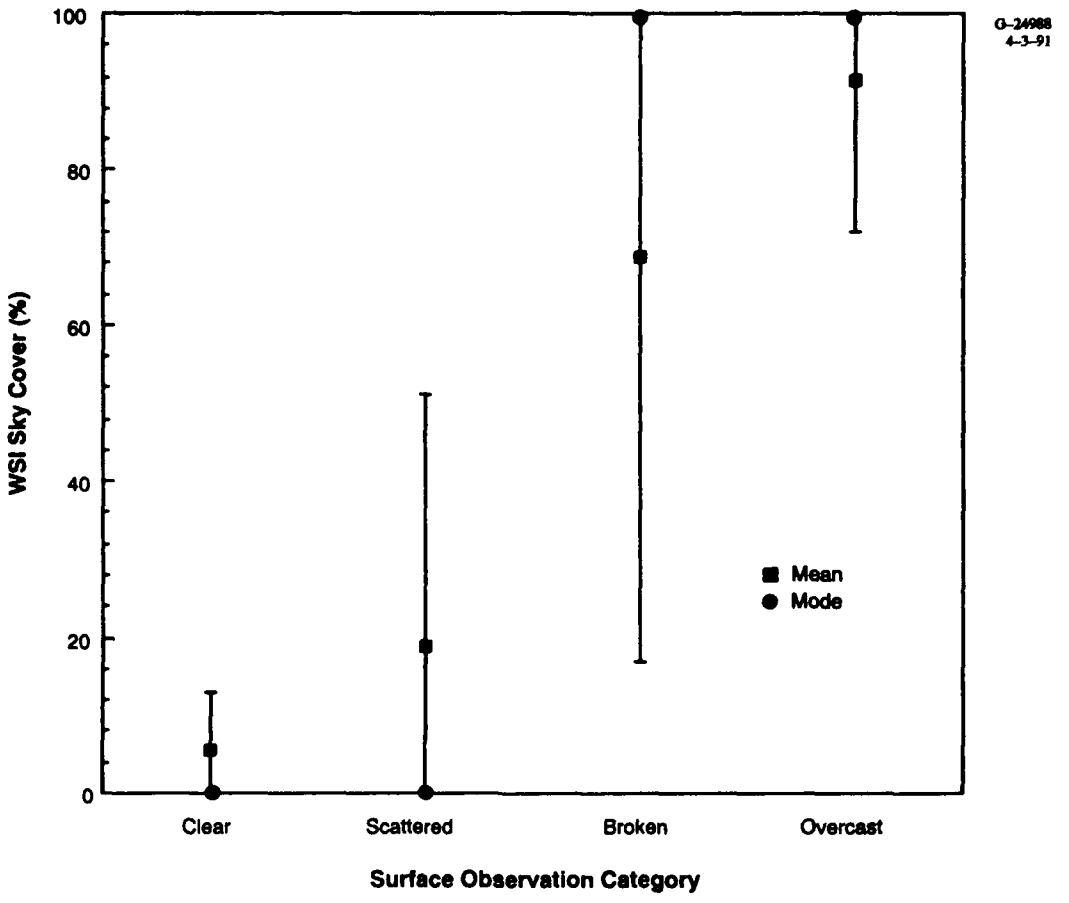


Figure 2-4 WSI/Surface Observation Comparison — Kirtland 10/90 Percentile Bounds

A portion of the actual distribution for WSI sky cover values for which the surface observation was clear is presented in Fig. 2-5a. Analogous distributions are shown, respectively, in Figs. 2-5b, 2-5c and 2-5d for the cases in which the surface observation was scattered, broken and overcast.

A combination of several reasons may explain the spread of the differences between WSI-derived and surface-observed sky cover:

- The subjective nature of estimating sky cover surely contributes to the spread seen in Figs. 2-3 and 2-4, especially for the scattered and broken categories. The human tendency to overestimate sky cover may account for many of the low values in the broken category.

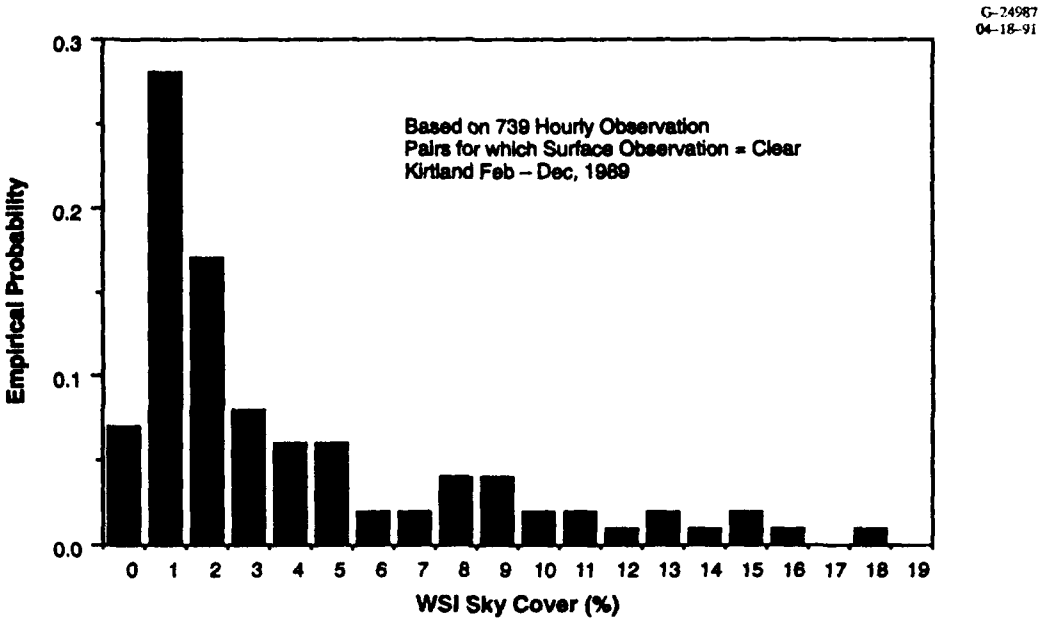


Figure 2-5a WSI Sky Cover for Clear Surface Observations

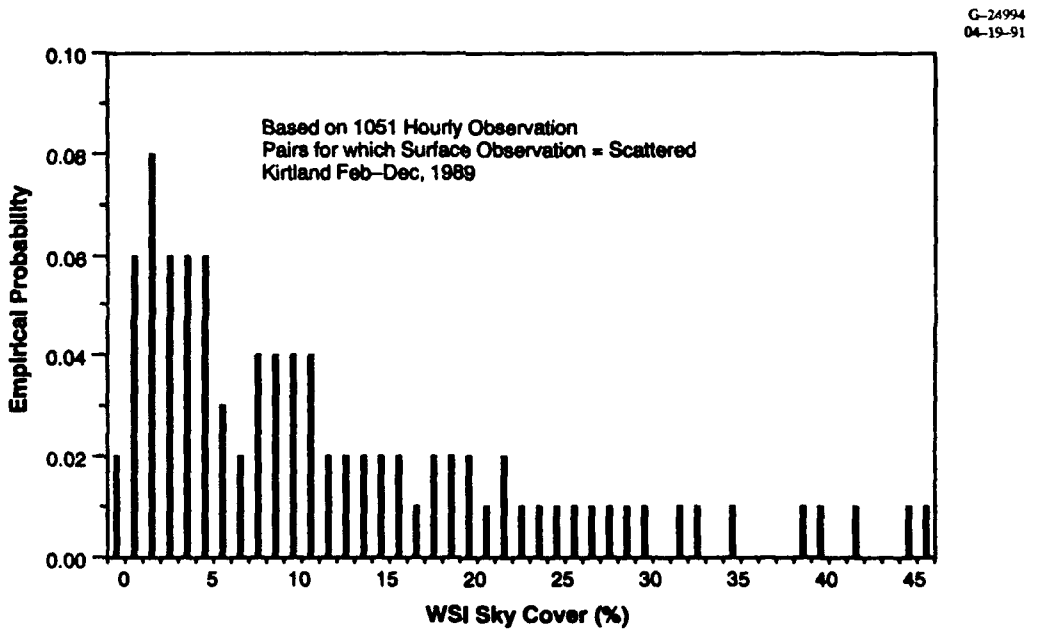


Figure 2-5b WSI Sky Cover for Scattered Surface Observations

WSI Sky Cover (%)

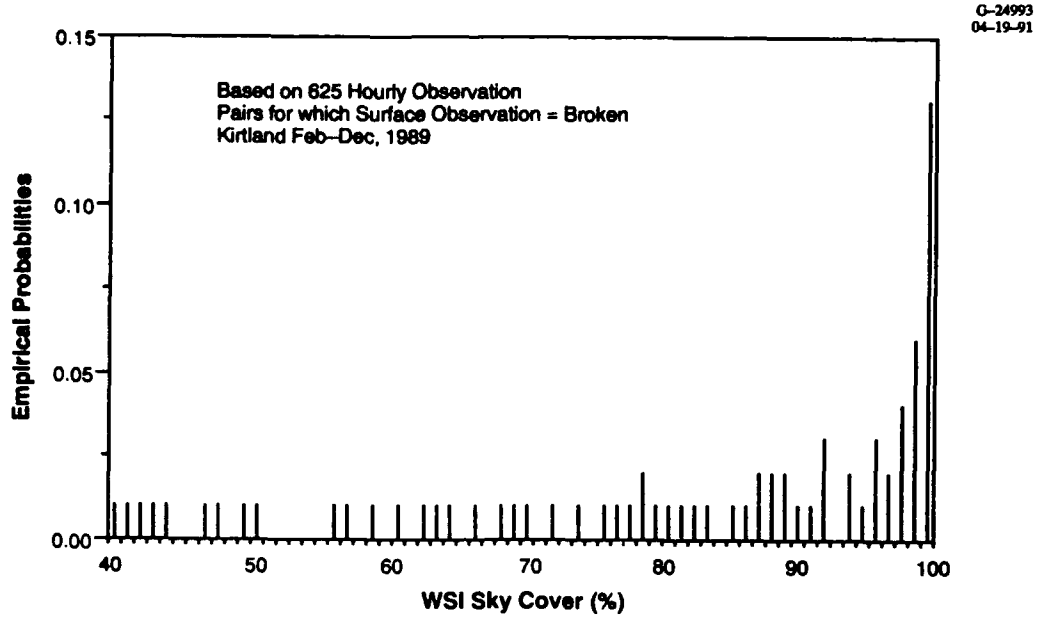


Figure 2-5c WSI Sky Cover for Broken Surface Observations

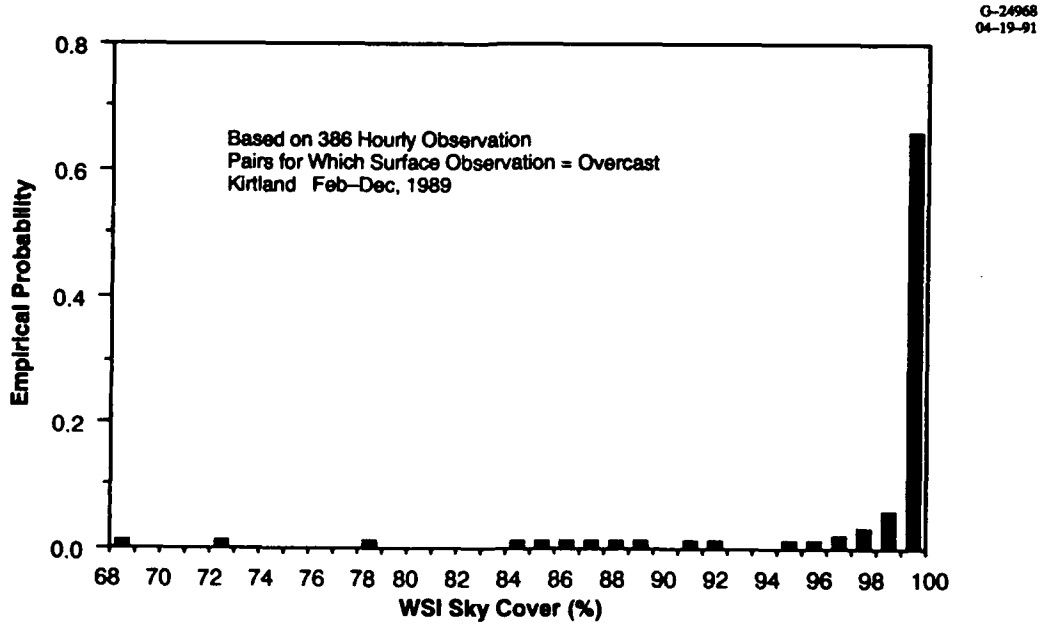


Figure 2-5d WSI Sky Cover for Overcast Surface Observations

- Erroneous indications of cloud around the occultor and close to the horizon may account for some of the higher values of WSI sky cover in the clear and scattered categories.
- It is possible that the Whole Sky Imager picks up haze and thin cloud more than a human observer.
- Under rapidly changing weather conditions, the precise time at which the surface observation is assumed to have been taken may be critical. As indicated above, ten minutes before the hour is our nominal assumption.

Procedures were implemented to automatically identify, by site, date, and time, those WSI images grossly inconsistent with the concurrent surface observations for possible subsequent manual inspection.

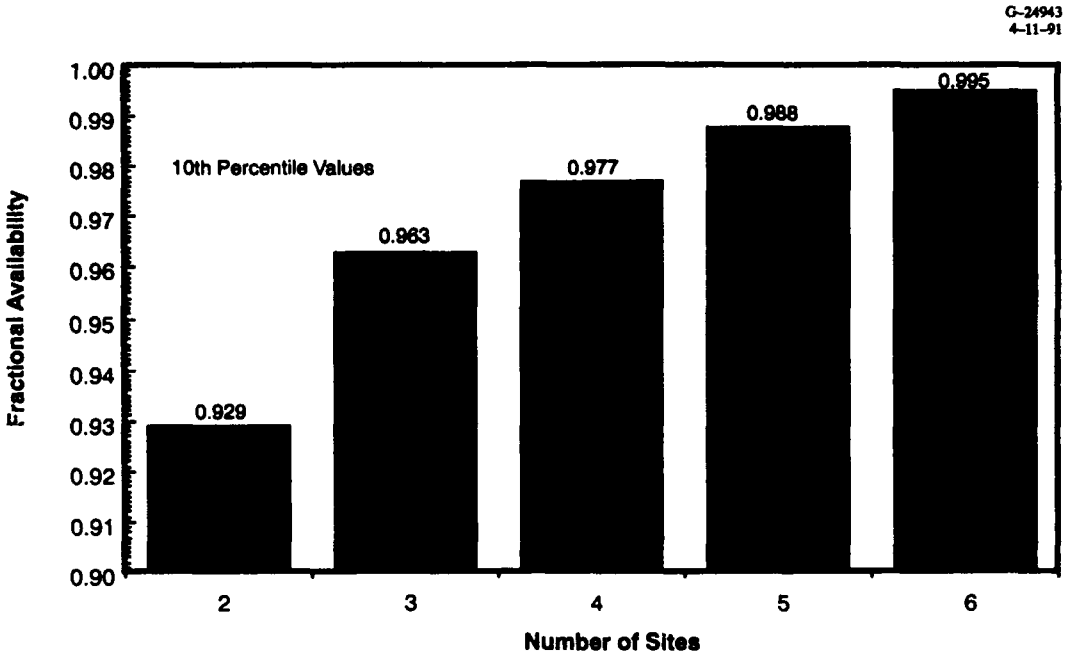
2.3.2 Impact of CFLOS4D Prediction Error

Substantial effort was devoted in the closing months of the project (September - December 1990) to examining the impact of cloud models on GBL site selection decisions. The intent was to indicate the benefit of collecting enough WSI data to perform reliable statistical validation analyses quantifying the accuracy of the model predictions so that a proper determination of site requirements is made. A meeting was held at the Pentagon on September 13 with LCDR J. Garner on this crucial issue.

Figure 2-6 protrays CFLOS4D system availability predictions for the sites and satellite location indicated in the figure. The values plotted are 10th percentile quantities: 90% of the one-year realizations simulated achieved higher availability. These CFLOS4D predictions are depicted in Fig. 2-7 as the small squares and represent true system availability as predicted by a perfect simulator. Ignoring the possibility of prediction errors and using these values directly results in the selection of the 3-site system given the postulated availability requirement of 95.5% (the dashed line).

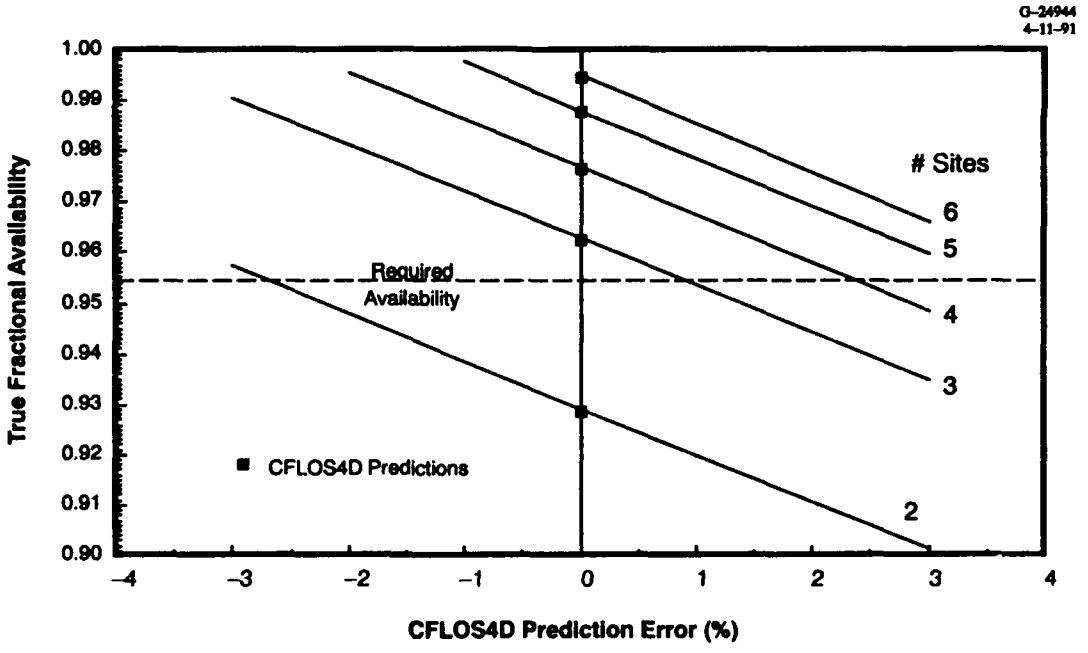
The oblique lines represent the true system availability corresponding to an incorrect CFLOS4D simulator which produced the values indicated by the squares. For example, if the fractional availability is overpredicted by 3%, then although CFLOS4D predicts 96.3% availability for three sites, the system would actually be available less than 94% of the time. If the actual prediction error were 3%, then it is clear from Fig. 2-7 that not only would the 3-site configuration fail to achieve the requirement and underperform, but the 4-site system would as well. Thus a 5-site system is needed although CFLOS4D predicted three.

Acknowledging the possibility of an availability prediction error as high as 3%, the 5-site configuration must be selected to ensure meeting the operational requirement. However, by way of validation



Order of Sites: White Sands, China Lake, Kirtland, Malmstrom, Columbia, Malabar
Geostationary Satellite Longitude: 112 deg W

Figure 2-6 CFLOS4D Availability Predictions



Order of Sites: White Sands, China Lake, Kirtland, Malmstrom, Columbia, Malabar
Geostationary Satellite Longitude: 112 deg

Figure 2-7 Significance of CFLOS4D Accuracy Assessment

analysis using a suitably-sized WSI image data set, if 2% accuracy can be assured (with any necessary model adjustments), then the 4-site system can be confidently selected, saving an unnecessary site. The point is that CFLOS4D predictions must be validated to ensure a proper determination of siting requirements.

3. SYSTEM LEVEL INVESTIGATIONS

Model validation data requirements and validation analyses are developed within the realm of statistical hypothesis testing. The concepts of statistical confidence, level of significance, and the power of a test are critical elements of the approach and serve as criteria in determining the degree to which we are assured of model validity. These validation criteria are discussed in this section. The approach for carrying out the data requirements and validation analyses, along with the assumptions made, are then presented and described. Results are discussed, with emphasis placed first on the validation accuracy that can be expected using the current WSI data set and second on the validation accuracy that could be obtained using a larger data set.

The CFLOS simulators provide downtime predictions on a simulated annual basis. Depending on the application of interest, worst case downtime results over a twenty year period, average yearly downtime, etc., may be typical predictions of interest. Year-to-year downtime distributions for multiple site cases, however, cannot be accurately determined from a few months of WSI data. Nonetheless, point comparisons of data-derived results with simulator-derived distributions assembled from hundreds of realizations over the length of the available data period can and have been made. The point comparisons, discussed in Section 3.1, serve primarily as a prelude to and motivation for the more difficult and precise *day-to-day* statistical distribution comparisons that follow in Sections 3.2 and 3.3.

For the present effort, TASC has conducted detailed system level analyses of four separate CFLOS4D validation cases. Four different combinations of WSI data collection sites, time periods (in 1989 only) and satellite subpoint locations are represented in the design of these cases (Table 3-1).

3.1 POINT COMPARISONS

Figure 3.1-1a presents a comparison of the WSI data-derived downtime count and the corresponding CFLOS4D downtime count distribution for the 1-5 minute downtime duration category. The results pertain to the specific three-site scenario called Case I in Table 3-1. Thus, from the WSI Case I data, 655 downtimes (cloud obscured time intervals) of duration 1-5 minutes were counted. The CFLOS4D distribution (actually a histogram) was determined from 2500 realizations of the WSI data period for the Case I scenario. A corresponding comparison of downtime percentages in the 1-5 minute

Table 3-1 System Level Validation Cases

CASE	SITES	SATELLITE SUBPOINT		DATE RANGE	SAMPLE SIZE	
		LONGITUDE	LATITUDE		DAYS	HOURS
I	WSC/Kirt/Col	100 W	50 N	2/1/89 — 12/31/89	242	1983
II	WSC/Kirt/Malm/Col	100 W	50 N	3/1/89 — 9/30/89	137	1273
III	WSC/Kirt/Col	145 W	50 N	2/1/89 — 12/31/89	235	1910
IV	WSC/Kirt/Malm/Col	145 W	50 N	3/1/89 — 9/30/89	125	1155

Key: WSC = White Sands C Station, New Mexico
Kirt = Kirtland AFB, New Mexico
Malm = Malmstrom AFB, Montana
Col = Columbia, Missouri

category is provided in Fig. 3.1-1b. Thus the 655 downtime durations in the 1-5 minute category accounted for 1.05% of the total time period. No obvious data/model inconsistency is apparent, but the possibility that the simulator underpredicts downtimes in this category is suggested.

A similar comparison of Case I data-derived and simulator-derived downtime count results is provided in Fig. 3.1-2a for the 6-30 minute downtime duration category. A corresponding comparison for downtime percentages in this category is shown in Fig. 3.1-2b. These comparisons provide strong evidence that the simulator overpredicts downtime durations in the 6-30 minute category. A similar tendency is further suggested, albeit less strongly, by the comparisons in Fig. 3.1-3a (downtime counts) and Fig. 3.1-3b (downtime percentages) for the 31-180 minute category, and in Fig. 3.1-4a (downtime counts) and Fig. 3.1-4b (downtime percentages) for the greater-than-three-hours category. However, data/model inconsistency is most dramatic in the 6-30 minute category.

Data-derived total percent downtime for the Case I scenario is compared with the simulator-derived distribution in Fig. 3.1-5. Total percent downtime for the WSI data sequence is simply the sum of the downtime percentages in each of the downtime duration categories. In aggregate, an overprediction of downtime appears to be a strong possibility.

In summary, these comparisons provide:

- marginal support to the suggestion that the simulator underpredicts downtime counts of short duration
- strong evidence of overpredicting downtime counts of longer duration, especially in the 6-30 minute range.

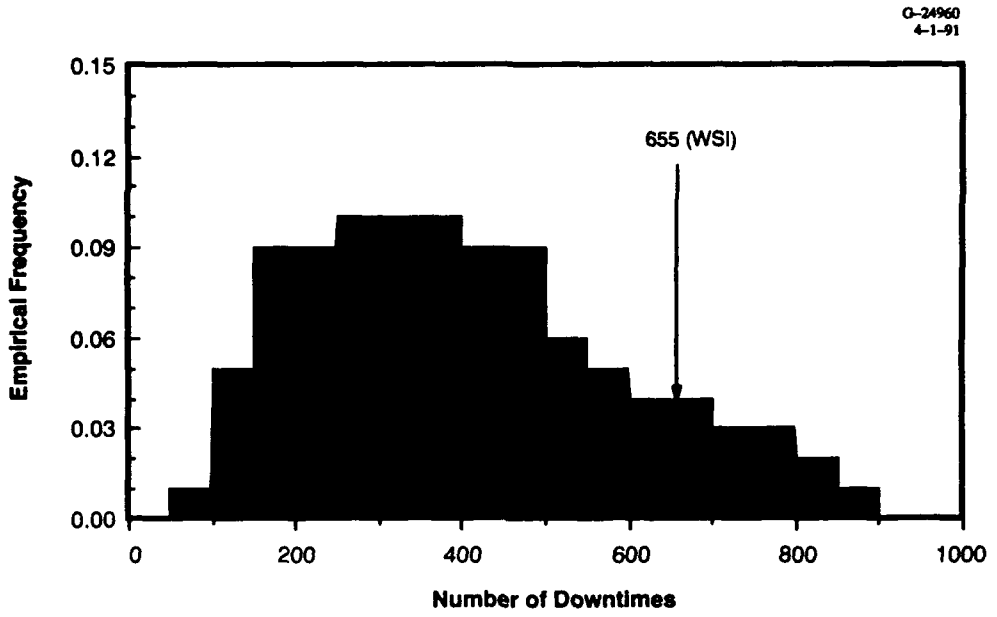


Figure 3.1-1a Case I CFLOS4D Downtime Count Distribution and WSI Data-Derived Count (Downtime Duration Category: 1-5 minutes)

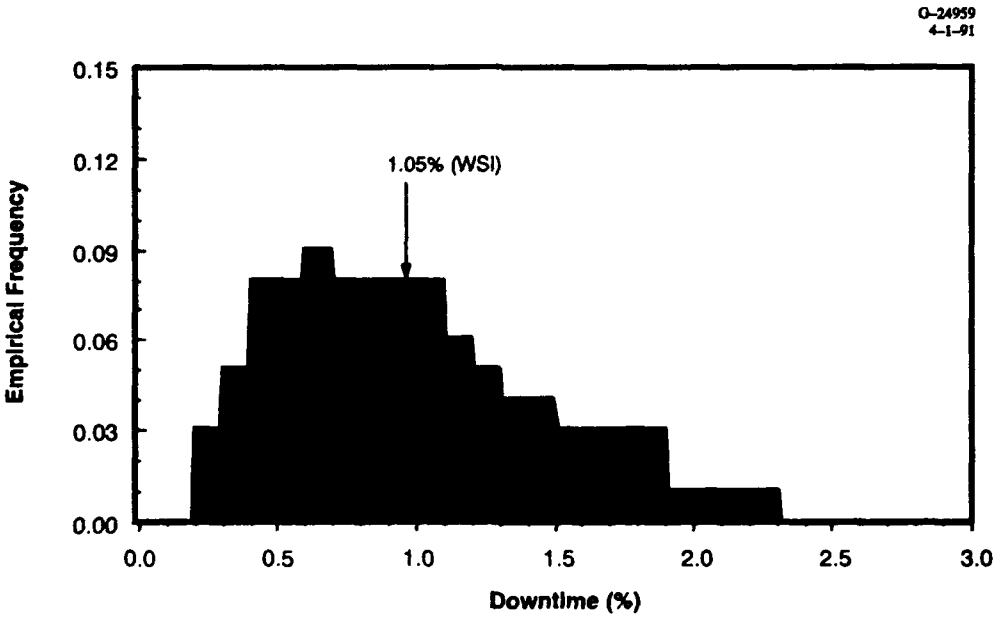


Figure 3.1-1b Case I CFLOS4D Downtime Percent Distribution and WSI Data-Derived Downtime (Downtime Duration Category: 1-5 minutes)

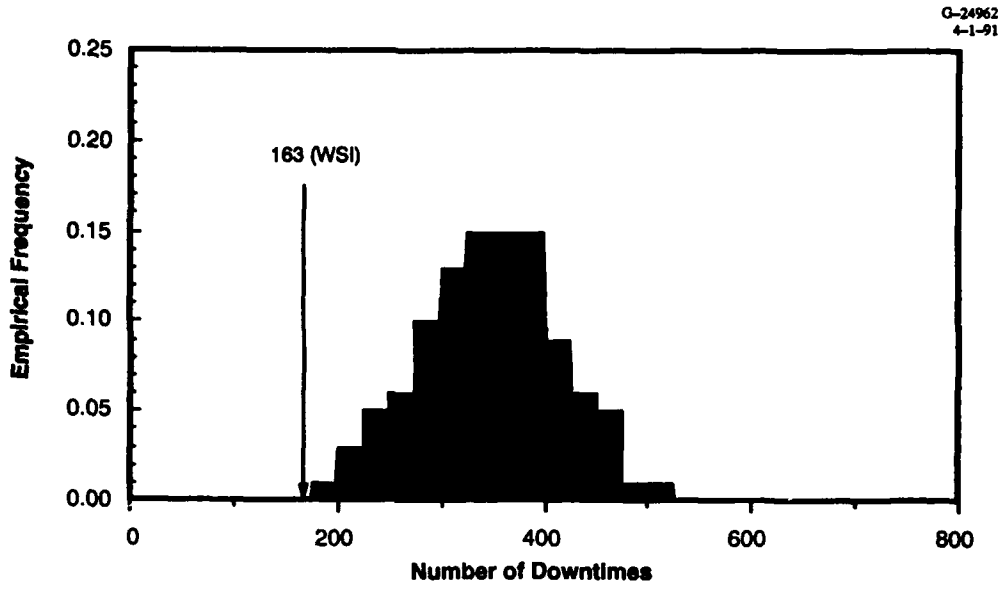


Figure 3.1-2a Case I CFLOS4D Downtime Count Distribution and WSI Data-Derived Count (Downtime Duration Category: 6-30 minutes)

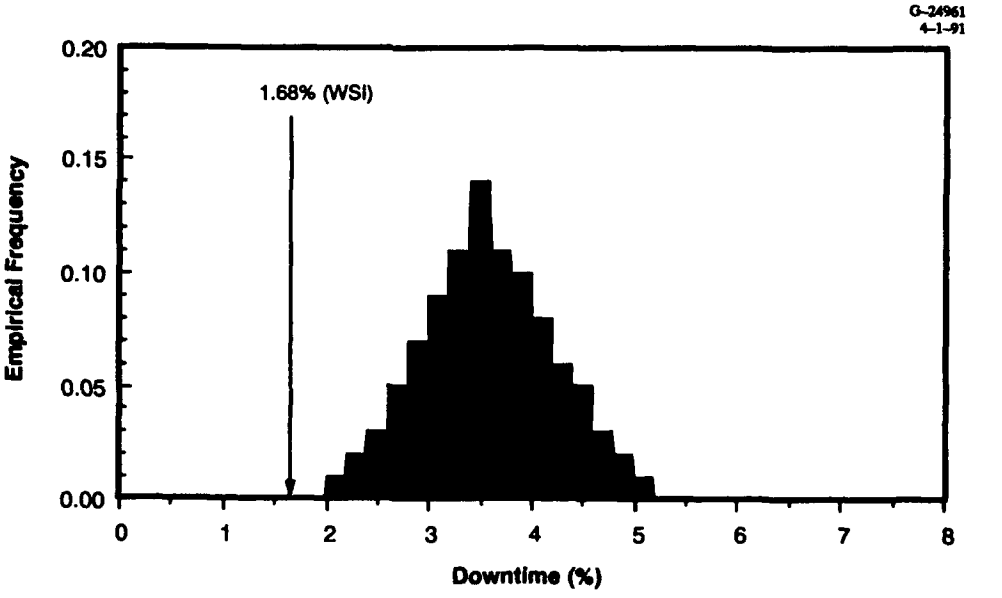


Figure 3.1-2b Case I CFLOS4D Downtime Percent Distribution and WSI Data-Derived Downtime (Downtime Duration Category: 6-30 minutes)

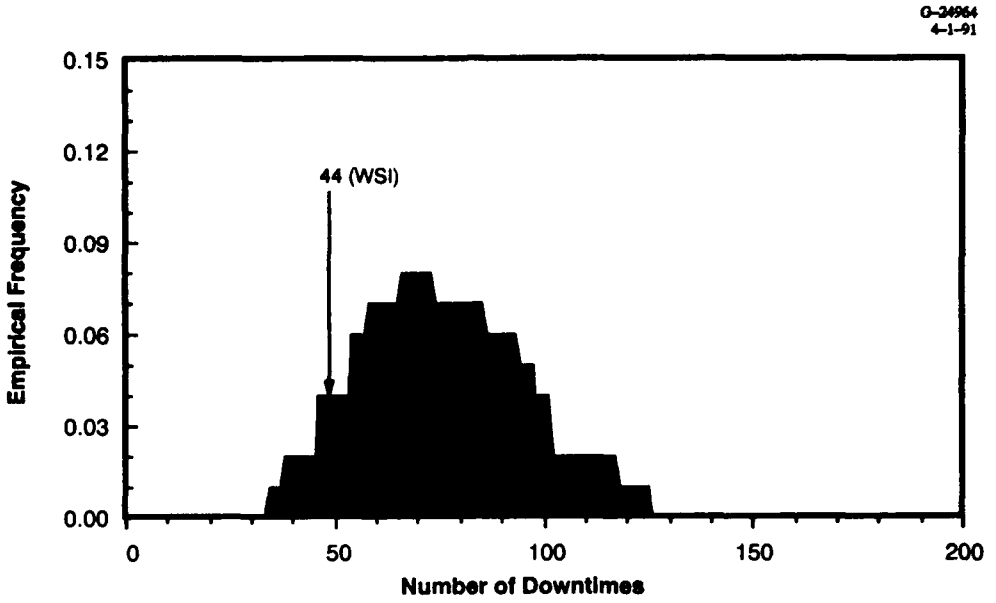


Figure 3.1-3a Case I CFLOS4D Downtime Count Distribution and WSI Data-Derived Count Downtime Duration Category: 31-180 minutes

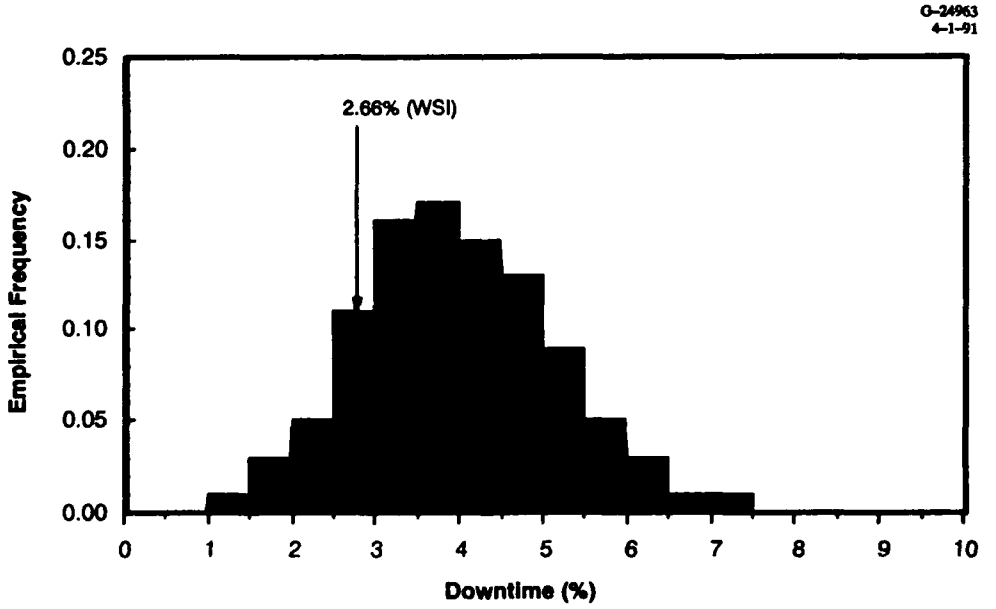


Figure 3.1-3b Case I CFLOS4D Downtime Percent Distribution and WSI Data-Derived Downtime Duration Category: 31-180 minutes

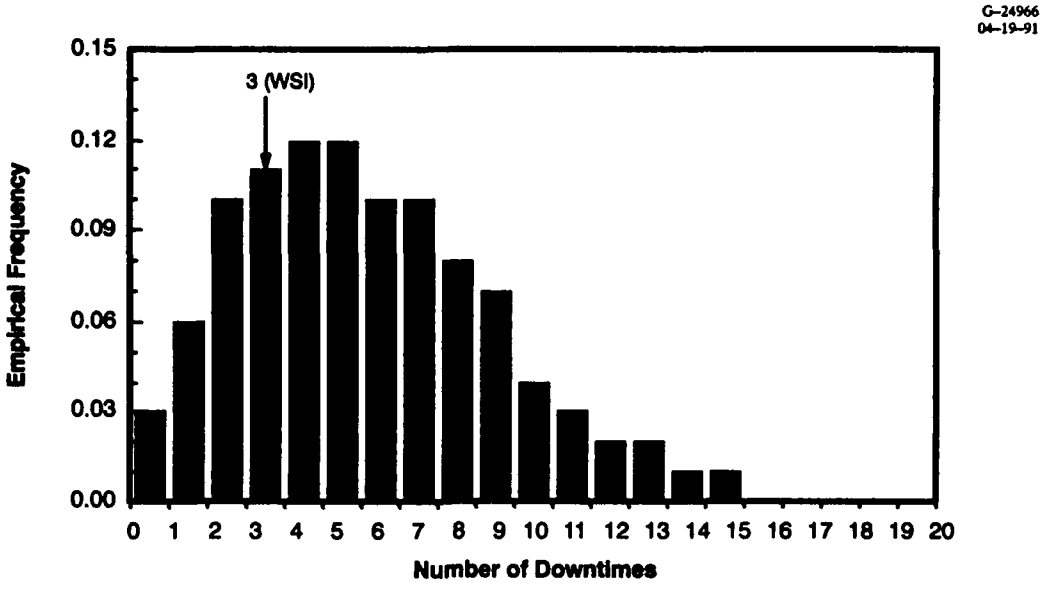


Figure 3.1-4a Case I CFLOS4D Downtime Count Distribution and WSI Data-Derived Count (Downtime Duration Category: Greater-than-Three-Hours)

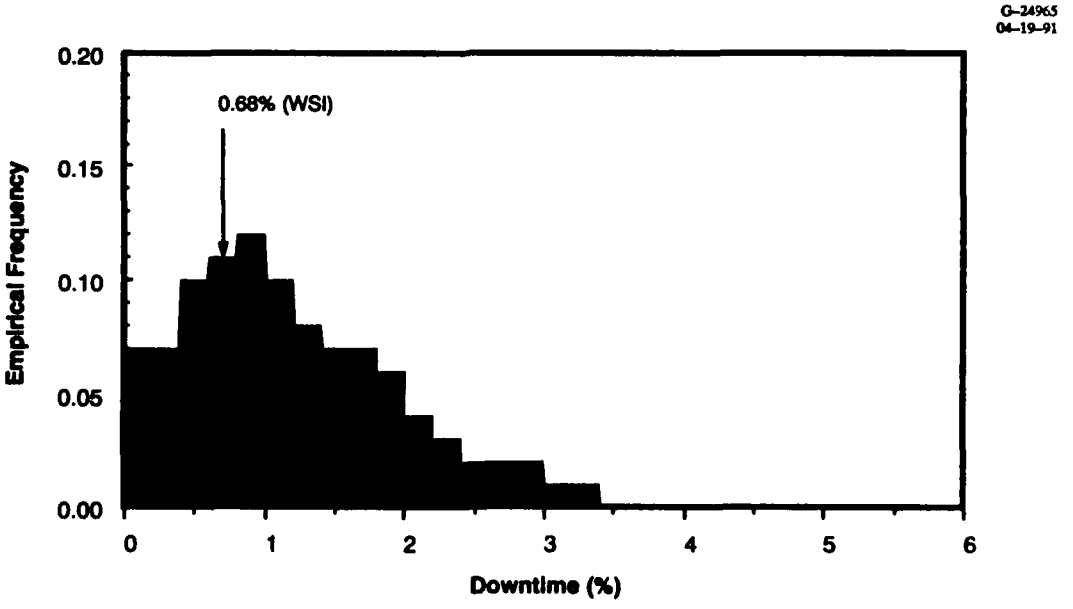


Figure 3.1-4b Case I CFLOS4D Downtime Percent Distribution and WSI Data-Derived Downtime (Downtime Duration Category: Greater-than-Three-Hours)

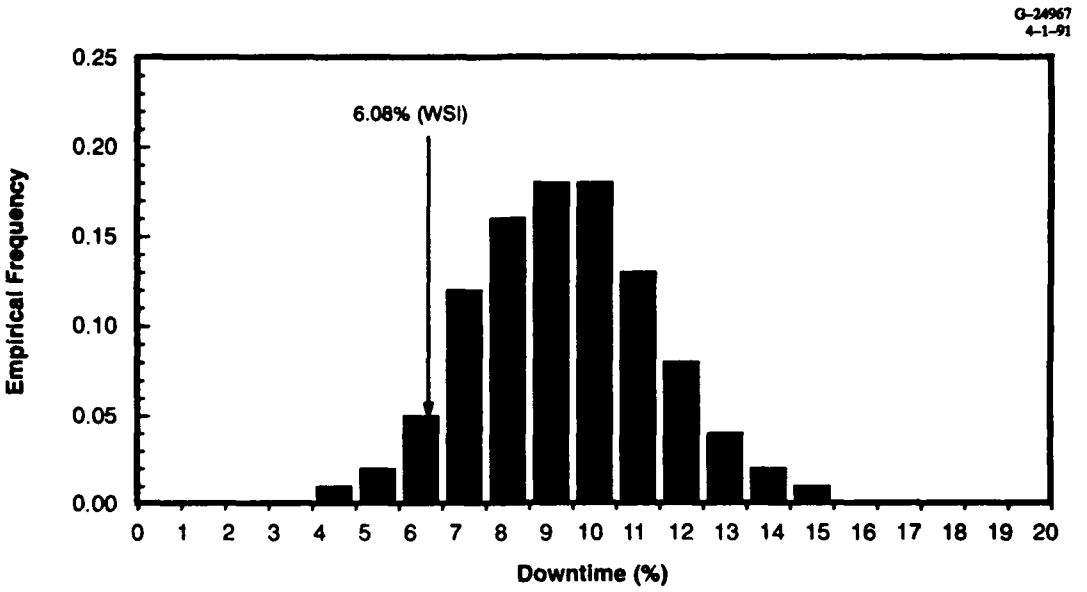


Figure 3.1-5 Case I CFLOS4D Total Percent Downtime Distribution and WSI Data-Derived Downtime

(What is unknown, of course, is the year-to-year variability of data-derived downtime counts; the full distribution of *day-to-day* data-derived downtime counts is exploited in the following sections.) An analogous comparison of results for a related four-site scenario (Case II) does nothing to alter this assessment, and Case II comparisons are therefore not provided. A plausible explanation for the tendency to underpredict short duration counts and overpredict long duration counts is that the modeled temporal correlation is too high. More will be said about this in Section 4.

3.2 CRITERIA FOR DAY-TO-DAY DISTRIBUTION COMPARISONS

In rigorously carrying out the test procedures we must acknowledge and account for the natural variability of the atmosphere in producing cloud patterns and motions which yield the cloud-free lines-of-sight that are of interest. We must note that our collection of observed cloud/no-cloud data is but one sample from an infinite population. Differences between observed and simulated data will surely result from sample variability; these differences have no bearing on model accuracy. We must take steps to avoid concluding that the simulations are invalid due to such differences. The criteria established in this section ensure that sample variability is accounted for in executing the validation procedure.

Our validation process must answer the question:

> Do the CFLOS4D simulations provide results that are consistent with observed WSI cloud/no-cloud data ?

This question is mathematically posed as a hypothesis to be tested. The central hypothesis, denoted by H_0, is

$$H_0\text{: Model and data are consistent.} \qquad (3.2\text{-}1)$$

One quantitative form of this hypothesis is presented below in Section 3.3.

The reliability of the hypothesis test is quantified by two factors, the confidence level, or equivalently, the *level of significance*, and the *power* of the test. It is the specification of these quantities, described below, which dictate the amount of observed data required for validation. (A third parameter, to be described shortly, must also be specified.) If conclusions about model validity are to be drawn with a high degree of reliability, a large amount of data may be required.

The *level of significance*, traditionally denoted by α , is the probability that the hypothesis is rejected when it should be accepted. This is also termed the false alarm rate. Expressed mathematically,

$$\alpha = \Pr \{ \text{ rejecting } H_0 \mid H_0 \text{ true } \}. \qquad (3.2\text{-}2)$$

In our case, if testing is performed at a level of significance of 0.1, then there is a 10% chance that we would conclude that the simulation is inconsistent with the observed data when in fact the opposite is true. This is called a Type I error. Equivalently, we are 90% *confident* that our conclusion is correct. That is, the confidence level, γ , is given by

$$\gamma = (1 - \alpha) \times 100\% . \qquad (3.2\text{-}3)$$

The *power* of a test is given by $1-\beta$, where β is the probability that the hypothesis is accepted when it should be rejected (a Type II error):

$$\beta = \Pr \{ \text{ accepting } H_0 \mid H_0 \text{ false } \}. \qquad (3.2\text{-}4)$$

Thus the power of the test is the probability of *correctly* detecting that the model is inconsistent with the observed data. Note that the conditioning event for the β probability definition (the condition that H_0 is false) is more complicated than the conditioning event for the α probability definition (the condition that H_0 is true). The magnitude of β depends on exactly how the hypothesis H_0 is not satisfied.

Another parameter involved in the criteria for model validation is one which defines the 'closeness' of the simulation results to observed data in the context of hypothesis testing. This parameter is termed the 'tolerance ratio' in the next section and is denoted by ψ. The parameter ψ is defined by a selected scalar measure of the maximum allowable deviation between model and data-derived distributions in accepting model/data consistency; in essence, ψ makes precise the intended meaning of "consistency" in Eq. 3.2-1. We shall utilize the parameter ψ in two different ways. In Section 3.4.1, when discussing validation results, ψ is prescribed in the design of the statistical test; in Section 3.4.2, when discussing data requirements analysis results, ψ is used instead to characterize the performance of the test.

In Section 3.4.2, the selected value of ψ greatly influences the amount of data needed for validation testing with specified reliability. On one hand, a value of $\psi = 1$ implies testing for strict equality of model and data distributions, an overly stringent condition. If a value of $\psi = 1.15$ is imposed while testing, then acceptance of the hypothesis H_0, that model and data are consistent, means that model predictions fall within 15% of reality as represented by the WSI data set. Note that the test procedure accounts for sample variability even with $\psi = 1$ since both α and β are nonzero. If we think of prediction error as being composed of both a random and a deterministic component, then testing with $\psi = 1$ incorporates only the random component whereas testing with $\psi = 1.15$ incorporates both the random component and a 15% deterministic error (explained more fully in Section 3.3). That is, $\psi = 1$ corresponds to testing for exact equality of distributions (with allowance for sampling error up to bounds implied by the choice of α and β); $\psi = 1.15$ allows for even greater discrepancy between the distributions (up to 15%). In this sense ψ is a tuning parameter in determining the data required to test for a specified degree of model precision or, conversely, the validation accuracy that can be achieved for a given amount of data available.

To instill faith in the conclusions of the hypothesis tests it is naturally desired that α, β and $|1 - \psi|$ all be small. However, these quantities cannot be made arbitrarily small without significantly impacting the amount of data required to test to such reliability. In the next section parametric equations illustrating the relationships among α, β and ψ and the corresponding amount of data available or required are presented.

3.3 APPROACH AND ASSUMPTIONS

We've adopted a nonparametric (distribution-free) approach to the validation hypothesis testing procedure. The random variables of interest here are 'downtime duration counts' per day, i.e., the number of times per day the system is down for a specified downtime duration category. (The downtime

categories are defined in Section 2.) Alternative choices of random variables are possible. One is 'downtime duration percentages' per day, i.e., the percentage of time per data-day the system is down for each specified downtime duration category. The histograms of such day-to-day downtime percentages do not appear to be significantly different from the corresponding day-to-day counts (this fact will be illustrated shortly) and thus validation results based on either choice would be similar.

Data requirements are initially determined to validate the 'single satellite' or single line-of-sight case, wherein only one simulation scenario with one satellite location, is defined. Thus, for CFLOS4D, for example, the amount of required data refers to the number of (12-hour) days of serial cloud/no-cloud data, from all sites, corresponding to the chosen satellite location (one pixel of data at each minute). However, by accounting for the effects of CFLOS sky dome spatial correlations, the time extent of data requirements can be reduced by including in the validation exercise several different CFLOS4D runs employing different line-of-sight directions (satellite locations). The degree of this reduction will be discussed at the end of Section 3.4.2.

Given a site/satellite configuration and a specified category of downtime durations (i.e., 1-5 minutes, 6-30 minutes, 31-180 minutes or greater-than-three-hours), suppose that in reality (as exemplified by the WSI data) downtime duration counts (per day)

$$X_1, X_2, \ldots, X_N \tag{3.3-1}$$

are from a population with cumulative distribution function F_{WSI}. (The procedure described has also been used for downtime duration time percentages and could be used for other choices of random variables as well.) Here X_i represents the downtime duration count for day i of the WSI data set and N represents the total number of available data-days. It is assumed that X_i and X_j are *statistically independent* for $i \neq j$. Note that the true distribution function F_{WSI} is unknown to the analyst but is approximated by the *empirical distribution function*

$$\hat{F}_{WSI}(x) = \frac{\text{the number of i such that } X_i \leq x}{N} \tag{3.3-2}$$

where x = 0, 1, 2, ... As N approaches infinity, $\hat{F}_{WSI}$ converges to F_{WSI} in probability.

The assumption of independence implies that day-to-day correlations of X_i values, the downtime duration counts per day, are zero. Day-to-day correlations would reduce the effective sample size, N_{eff}, (i.e., $N_{eff} = N$ if the downtime duration counts are uncorrelated, $N_{eff} < N$, otherwise) as determined in Ref. 7. To test with prescribed confidence, more data are required if correlations exist. A procedure

(based on the chi-squared test) to determine the degree of day-to-day dependence for both observed data and CFLOS simulator realizations has been developed, along with a procedure for computing effective sample size. An investigation based on both CFLOS4D simulation data and WSI data indicates that day-to-day downtime duration count correlations are insignificant for both single and multiple site cases.

Suppose also that multiple runs of the CFLOS4D simulator yield an empirical distribution function which we further assume is equal to the *true* (as opposed to *estimated*) simulator distribution function F_{SIM}. This is reasonable if the number of simulator runs is sufficiently large, e.g., greater than 500 one-year runs, and enables us to use a *one-sample* procedure for validation, specifically, the Kolmogorov-Smirnov (KS) goodness-of-fit test recommended by a number of sources (Refs. 11 and 12). By 'two-sample approach' we mean testing the consistency of two data sets; by 'one-sample approach' we mean testing the consistency of a single data set against a hypothesized truth.

In testing the truth of the null hypothesis represented by Eq. 3.2-1 and now quantified by

$$H_0 : F_{WSI} = F_{SIM} \tag{3.3-3}$$

one cannot use the standard tables for KS statistic α probabilities and test thresholds since F_{WSI} and F_{SIM} are discrete distribution functions. Monte Carlo simulation must be used to determine α and β probabilities and test thresholds, t, because no analytic formulations, applicable for potentially large N, are known (Refs. 11 and 12). One should keep in mind too that the tolerance parameter ψ plays a role in the definition of equality in Eq. 3.3-3. The way in which ψ enters the hypothesis test design will be clarified shortly.

The KS statistic is defined by

$$D = \max_{x = 0,1,2,\ldots} \sqrt{N} \left| F_{SIM}(x) - \hat{F}_{WSI}(x) \right|, \tag{3.3-4}$$

i.e., D is proportional to the maximum (over x, where x is a downtime duration count value) of the absolute differences between the (true) simulator-derived distribution function values and the (empirical) WSI-derived distribution function values (based on N days of data). Under H_0, test thresholds, t, are defined via

$$\Pr \{D > t\} = \alpha \tag{3.3-5}$$

for prescribed α and ψ. The actual validation test consists of rejecting H_0 if the observed value of the KS statistic D exceeds t and accepting H_0 otherwise.

The tolerance parameter ψ, introduced earlier in Section 3.2, quantifies the 'maximum allowable deviation' between the distributions F_{WSI} and F_{SIM} and is prescribed by the analyst prior to validation testing. Many possible definitions for ψ exist. One could, for example, define ψ as a 'normalized' KS statistic (Eq. 3.3-4):

$$\psi = \max_{x = 0,1,2,\ldots} \frac{| F_{SIM}(x) - F_{WSI}(x) |}{F_{WSI}(x)} . \tag{3.3-6}$$

Alternatively, one could define ψ via moments associated with F_{WSI} and F_{SIM}, for example,

$$\psi = \frac{\mu_{WSI}}{\mu_{SIM}}, \tag{3.3-7}$$

the ratio of the mean WSI day-to-day downtime duration count μ_{WSI} to the mean simulator day-to-day downtime duration count μ_{SIM}. We use Eq. 3.3-7 since, of all available statistical parameters, the mean downtime duration count appears to be the most important to GBL system designers. (Other definitions of ψ are certainly possible. Should it turn out that an alternative definition for ψ more closely reflects primary concerns regarding the differences between $F_{WSI}(x)$ and $F_{SIM}(x)$ in a future application, the validation approach is flexible and will accept alternative definitions.)

Two algorithms are now presented which are crucial to our analysis. The first, an algorithm for computing test thresholds, is closely linked to the actual validation procedure. It also underlies the second, an algorithm for computing false acceptance (β) probabilities, which in turn constitutes the foundation of the data requirements analysis procedure. Using these two algorithms, we can rigorously address the following two questions:

> In a test for model/data consistency with specified confidence level, tolerance ratio and (necessarily limited) data sample size, what is the corresponding test power?
>
> In a test for model/data consistency with specified confidence level, test power and tolerance ratio, what is the corresponding minimum required sample size?

Test powers ($1 - \beta$) and required sample sizes N, which characterize validation test performance over a range of significance levels α and tolerance ratios ψ, are summarized for this effort in Section 3.4.

3.3.1 The Test Threshold Algorithm

Figure 3.3-1 illustrates the algorithm for computing the test threshold, t, which, for a desired tolerance ratio $\psi = 1$ (the most stringent possible condition), depends on:

- the (true) distribution function F_{SIM} for CFLOS4D day-to-day downtime duration counts (set equal to simulator data results)

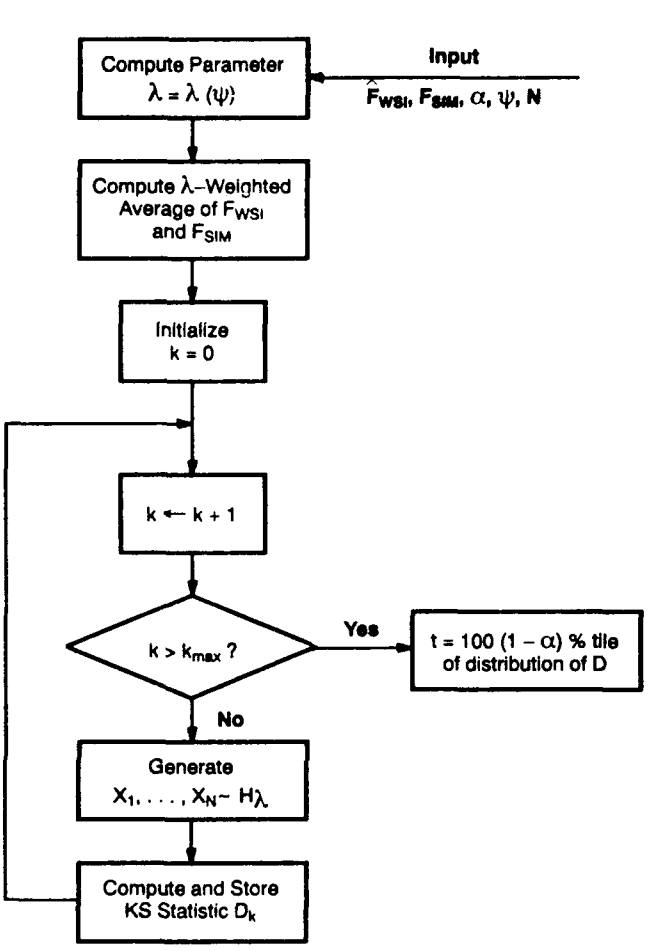


Figure 3.3-1 Algorithm for Computing Test Threshold t

- the chosen α probability
- the available WSI sample size N or a selected candidate sample size.

If $\psi \neq 1$ is desired (i.e., for allowable deviations of $|1 - \psi| * 100\%$), then the test threshold t depends on all of the above and, additionally, on

- the empirical distribution function $\hat{F}_{WSI}$ for WSI day-to-day downtime duration counts (from WSI data results).

Following these specifications, the algorithm

(i) computes a mixture distribution equal to the weighted average of $\hat{F}_{WSI}$ and F_{SIM}

$$H_\lambda(x) = (1-\lambda)\,\hat{F}_{WSI}(x) + \lambda F_{SIM}(x) \qquad (3.3\text{-}8)$$

where

$$\lambda = \lambda(\psi) = \frac{\psi - \frac{\hat{\mu}_{WSI}}{\mu_{SIM}}}{1 - \frac{\hat{\mu}_{WSI}}{\mu_{SIM}}} \qquad (3.3\text{-}9)$$

(note that $\lambda = 1$ if and only if $\psi = 1$, so H_λ is independent of $\hat{F}_{WSI}$ if and only if $\psi = 1$) and where $\hat{\mu}_{WSI}$ and μ_{SIM} are computed, respectively, from the WSI data and from CFLOS simulated data

(ii) randomly generates a 'synthetic' sequence $X_1, \ldots, X_N$ of WSI day-to-day downtime duration counts distributed according to H_λ

(iii) computes from this sequence the corresponding empirical distribution function $\hat{H}_\lambda$

(iv) computes the KS statistic D (Eq. 3.3-4) with $\hat{F}_{WSI}$ replaced by $\hat{H}_\lambda$

(v) repeats steps (ii) - (iv) as many times as desired to obtain an estimate of the distribution of D

(vi) defines t to be the 100 $(1 - \alpha)$ percentile of the distribution of D.

The weighted average distribution $H_\lambda(x)$ is the theoretical artifact which allows one to ascertain test thresholds for various ψ. When $\psi = \hat{\mu}_{WSI}/\mu_{SIM}$ then $H_\lambda = \hat{F}_{WSI}$; when $\psi = 1$ then $H_\lambda = F_{SIM}$; intermediate ψ values give rise to distributions H_λ intermediate to F_{WSI} and F_{SIM}.

3.3.2 The Test Power Algorithm

Figure 3.3-2 illustrates the algorithm for computing false acceptance (β) probabilities, given a validation test threshold t (computed, perhaps, as above). Here, we wish to determine the statistical power of the KS test to distinguish between two distributions F_{WSI} and F_{SIM}. To do this, the algorithm accepts as inputs:

- the (true) distribution function F_{SIM} for CFLOS4D day-to-day downtime duration counts (set equal to simulator data results)
- the available WSI sample size, N, or a selected candidate sample size.
- a hypothesized distribution function F_{WSI} for WSI day-to-day downtime duration counts (possibly $\hat{F}_{WSI}$ as given in Section 3.3.1)
- the test threshold t.

Following these specifications, the algorithm

(a) randomly generates a 'synthetic' sequence $X_1, \ldots, X_N$ of WSI day-to-day downtime duration counts distributed according to F_{WSI}

(b) computes from this sequence the corresponding synthetic empirical distribution function $\hat{F}_{WSI}$

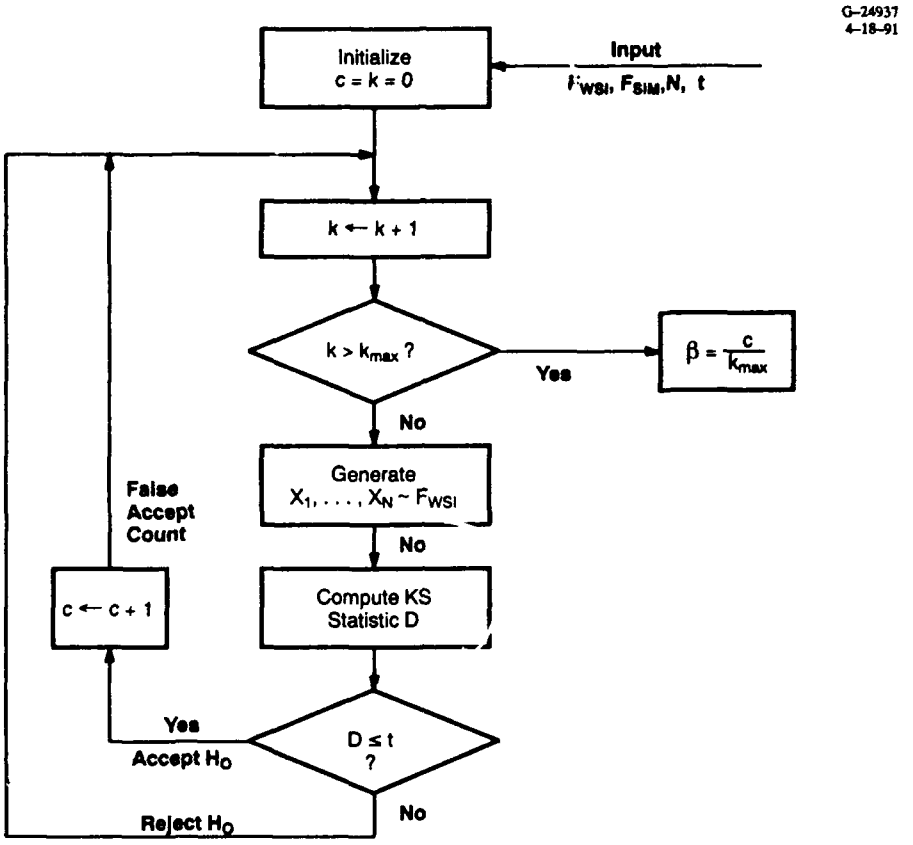


Figure 3.3-2 Algorithm for Computing β Probabilitities

(c) computes the KS statistic D (Eq. 3.3-4)

(d) repeats steps (a) - (c) as many times as desired, keeping track of how often H_0 is accepted, i.e., how often the KS statistic D is less than t.

The final estimate of β is obtained by dividing the false acceptance count by the total count of iterations.

3.4 RESULTS OF DISTRIBUTION COMPARISONS

We have conducted CFLOS4D system level validation analyses for the four cases listed in Table 3-1. Results for the four cases are quite similar, thus emphasis will be placed on Case I results in Section 3.4.1 in the interests of brevity. Associated data requirements results are exhibited in Section 3.4.2.

3.4.1 Validation Analyses

To assemble the simulator-derived distribution functions F_{SIM}, downtime duration counts for each category (1-5 minute, 6-30 minute, 31-180 minute and greater-than-three-hours) were compiled from 1000 CFLOS4D realizations for each of the four cases in Table 3-1. For consistency with the WSI data set, only results corresponding to the WSI data time periods (Table 3-1) were compiled.

Figures 3.4-1a to 3.4-4a exhibit the Case I CFLOS4D and WSI histograms of day-to-day downtime duration counts, respectively, for each of the four downtime categories. The differences between CFLOS4D histograms and WSI histograms are clearly visible for the 1-5 minute, 6-30 minute and 31-180 minute categories. The CFLOS4D simulator evidently *underestimates* (relative to the WSI data) the frequency of days for which *no* downtime (of the specified duration category) is accumulated, for these three categories. (More subtle distributional properties are responsible for the relative sizes of the *mean* downtime counts, also indicated in the figures.) Proper characterization of the distribution of the greater-than-three-hour category counts is especially vital, given their strategic importance as discussed in Section 2. More will be said about these particular category counts in Section 3.4.2 regarding data requirements.

Figures 3.4-1b to 3.4-4b exhibit the Case I CFLOS4D and WSI histograms of day-to-day downtime duration percentages, respectively, for each of the four downtime categories. Probabilities of greater than 0% downtime appear to be more widely spread than the corresponding probabilities of greater than 0 downtime counts. There is a somewhat arbitrary but unavoidable element of choice in the histogram bins in Figs. 3.4-1b to 3.4-4b (0.2%, 1%, 5% and 25% of day time, respectively). A different choice of histogram bins will, of course, alter the appearance of the histograms and possibly change, to some degree, the interpretation of the validation results. For this reason we shall present most of our analysis results in the downtime duration count domain and shall indicate time percentage results occasionally for the sake of comparison only.

Figures 3.4-5 to 3.4-8 exhibit the Case II CFLOS4D and WSI histograms of day-to-day downtime duration counts/percentages in a manner similar to Figs. 3.4-1 to 3.4-4. Once again the CFLOS4D simulator underestimates the probability of zero downtime counts. Downtimes are, not surprisingly, rarer for Case II than for Case I since Case II has one additional WSI site included. Histograms for Cases III and IV have similar characteristics to those in Figs. 3.4-1 to 3.4-8.

Table 3.4-1 provides estimates of the data/model mean downtime count ratios, $\dot{\psi}$, for each of the four cases and each downtime category. The $\dot{\psi}$ values are interesting statistics which give some

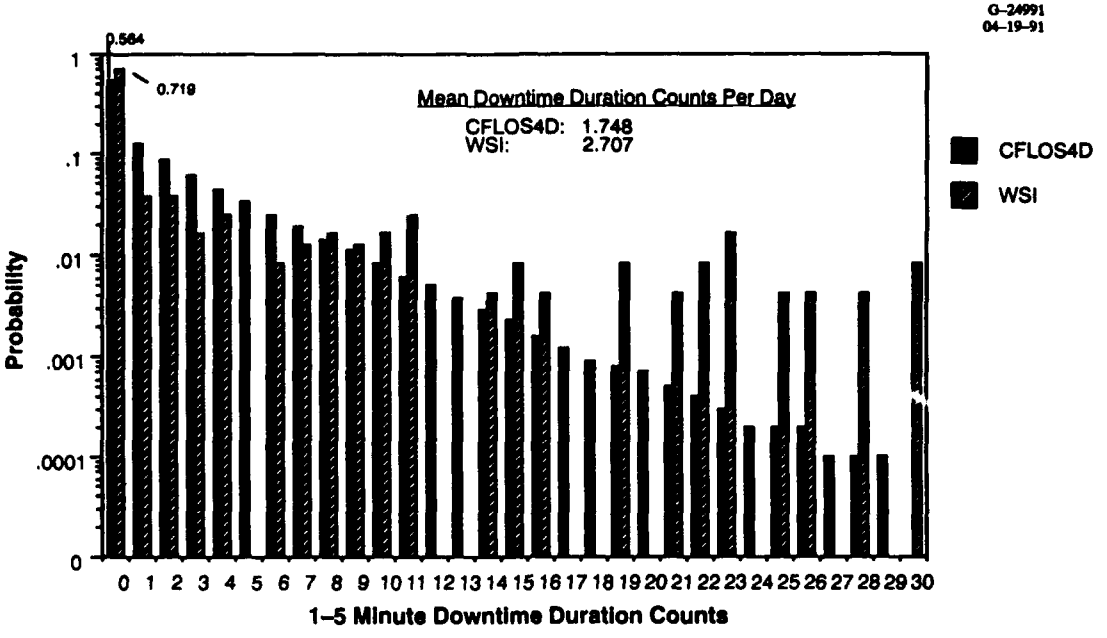


Figure 3.4-1a Case I Histogram for 1-5 Minute Category Counts

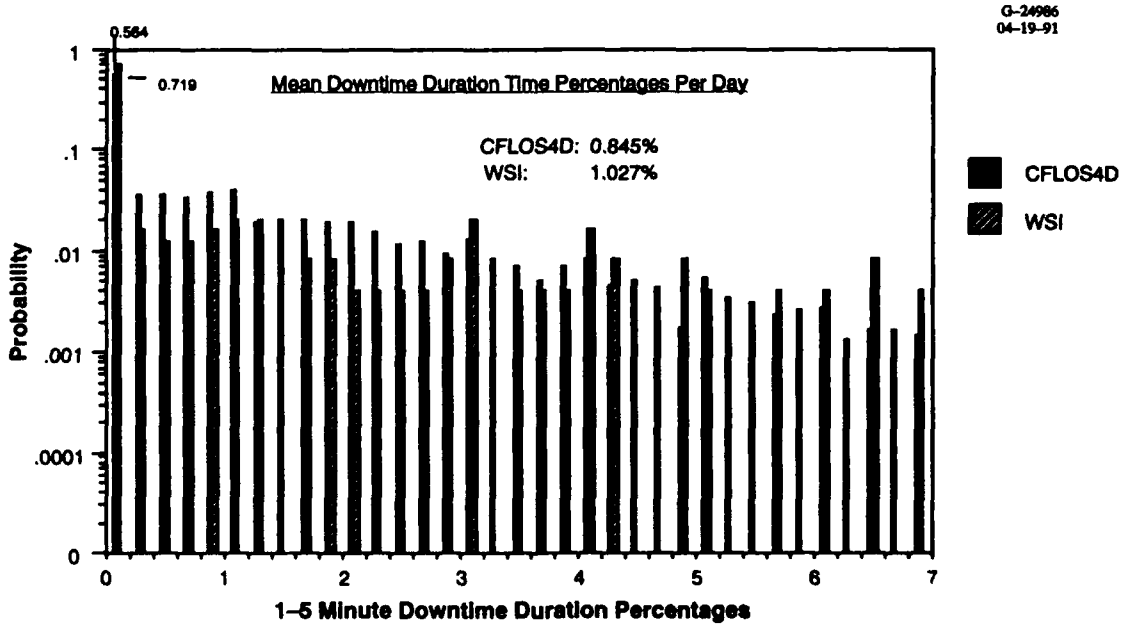


Figure 3.4-1b Case I Histogram for 1-5 Minute Category Percentages

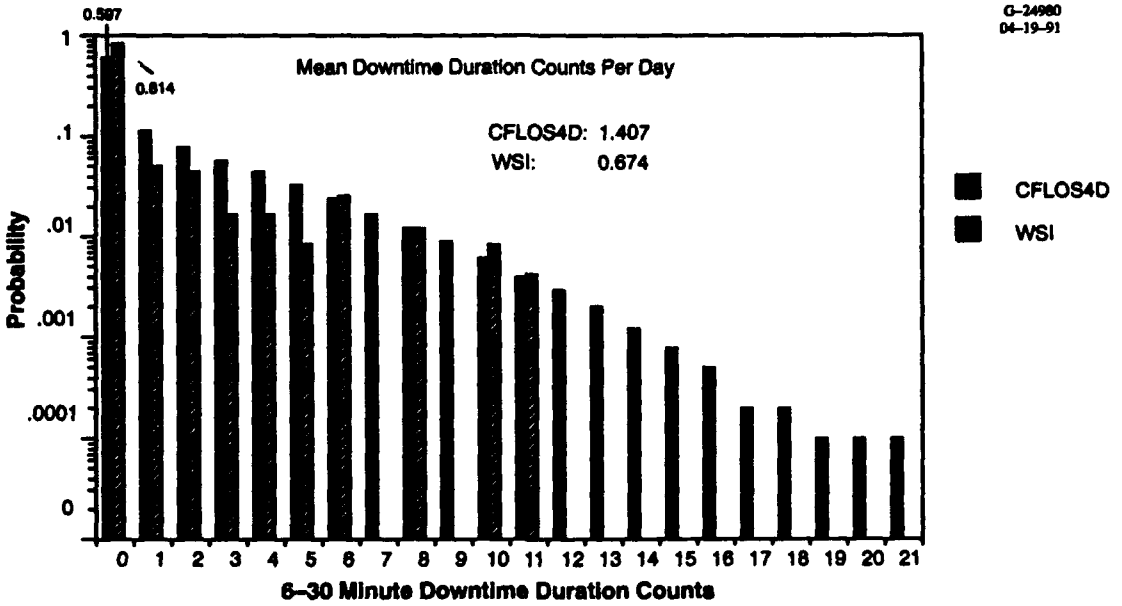


Figure 3.4-2a Case I Histogram for 6-30 Minute Category Counts

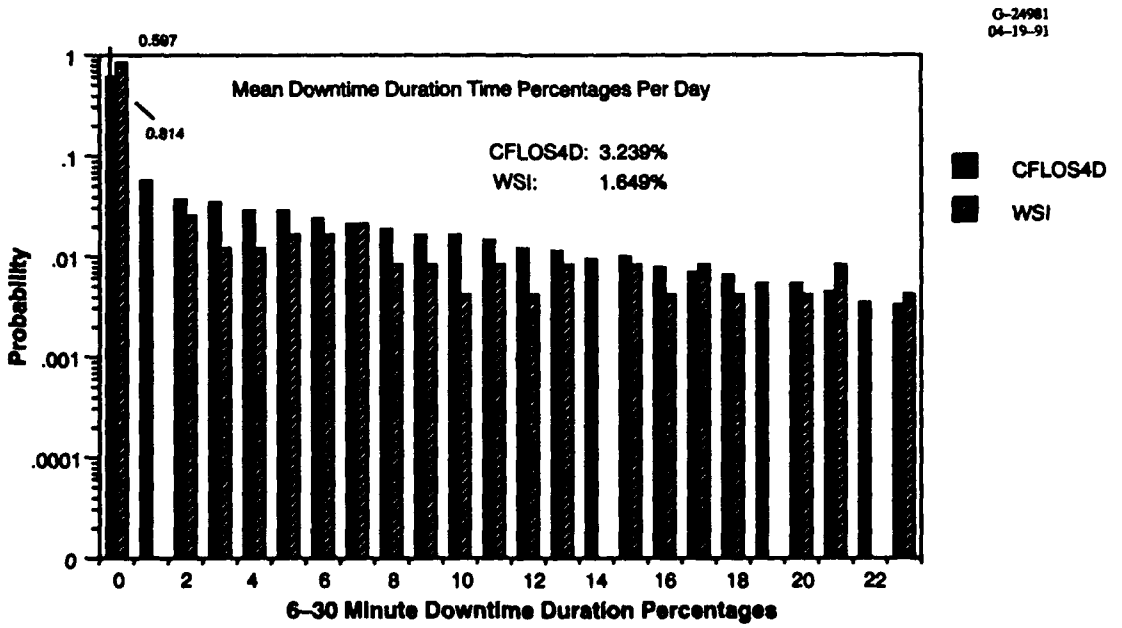


Figure 3.4-2b Case I Histogram for 6-30 Minute Category Percentages

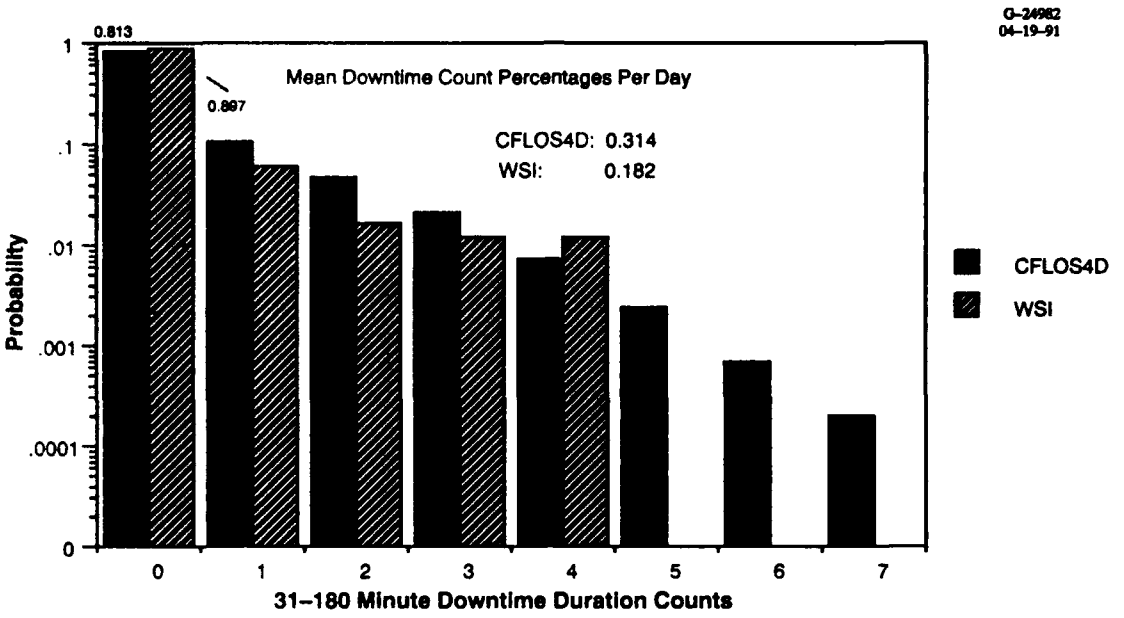


Figure 3.4-3a Case I Histogram for 31-180 Minute Category Counts

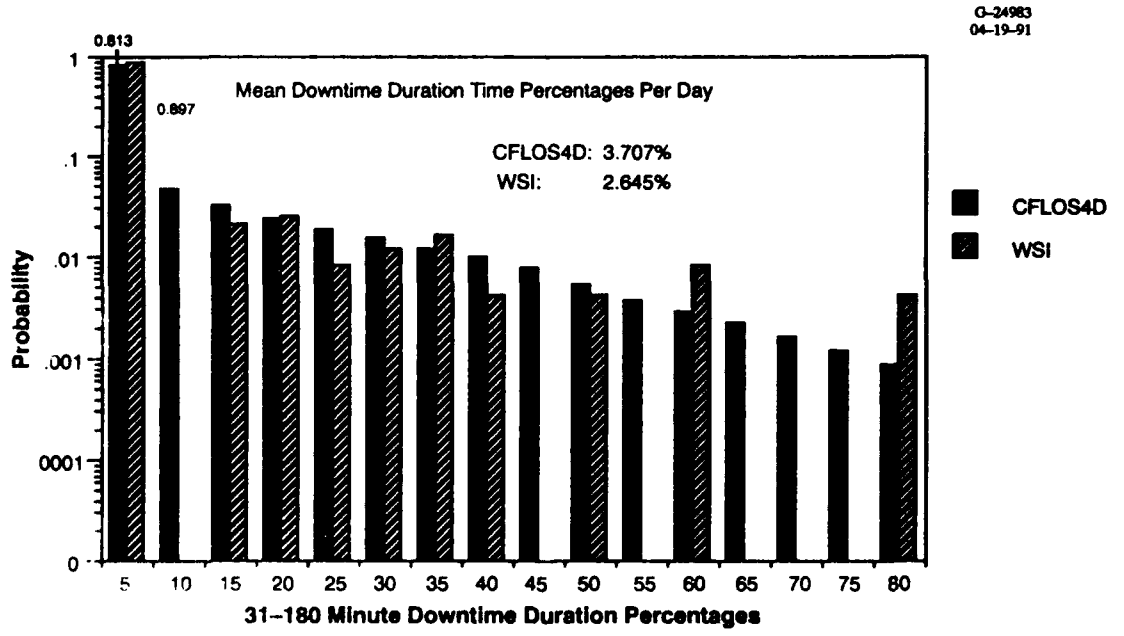


Figure 3.4-3b Case I Histogram for 31-180 Minute Category Percentages

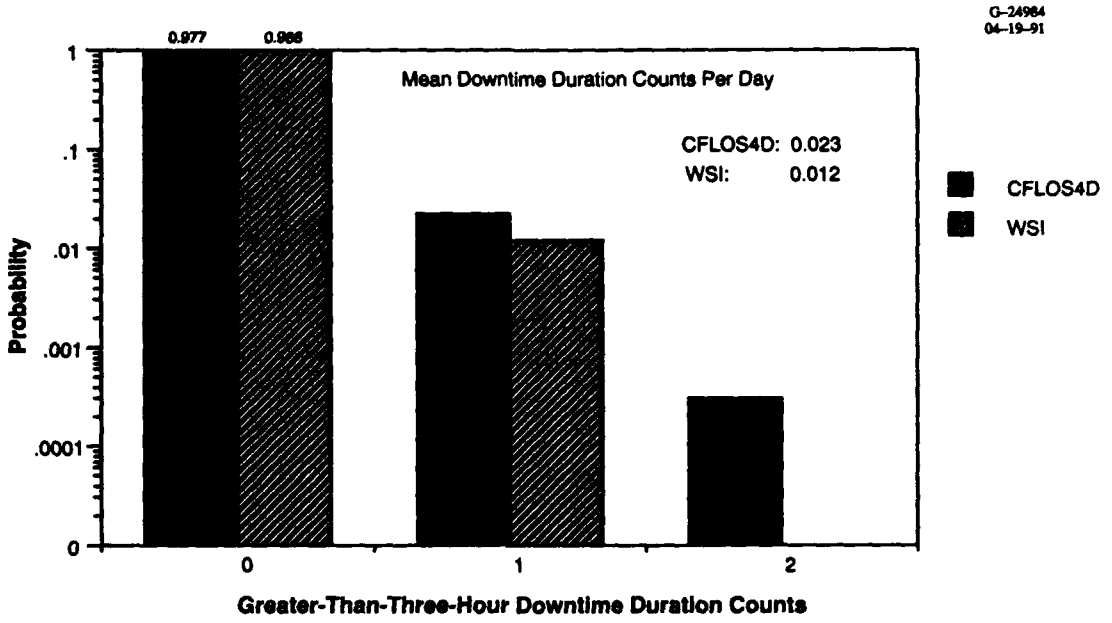


Figure 3.4-4a Case I Histogram for Greater-than-Three-Hour Category Counts

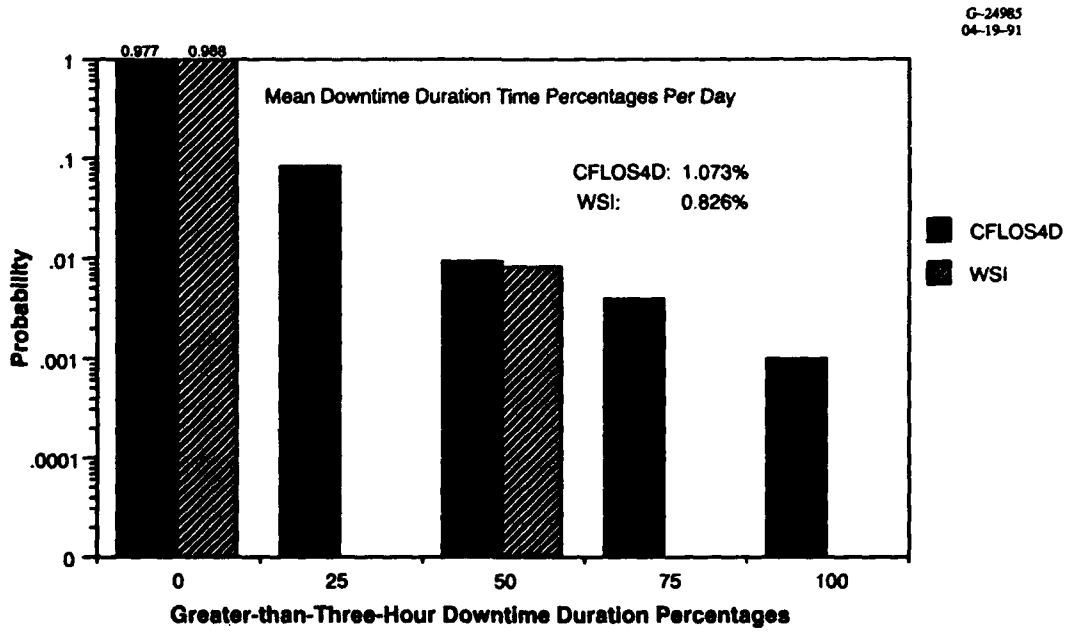


Figure 3.4-4b Case I Histogram for Greater-than-Three-Hour Category Percentages

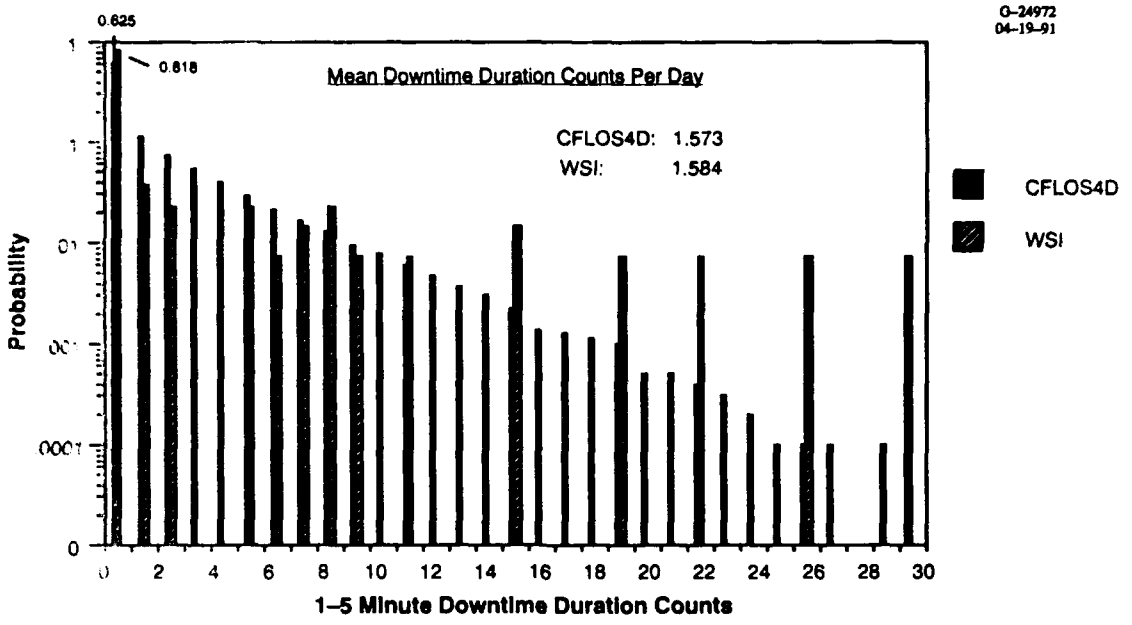


Figure 3.4-5a Case II Histogram for 1-5 Minute Category Counts

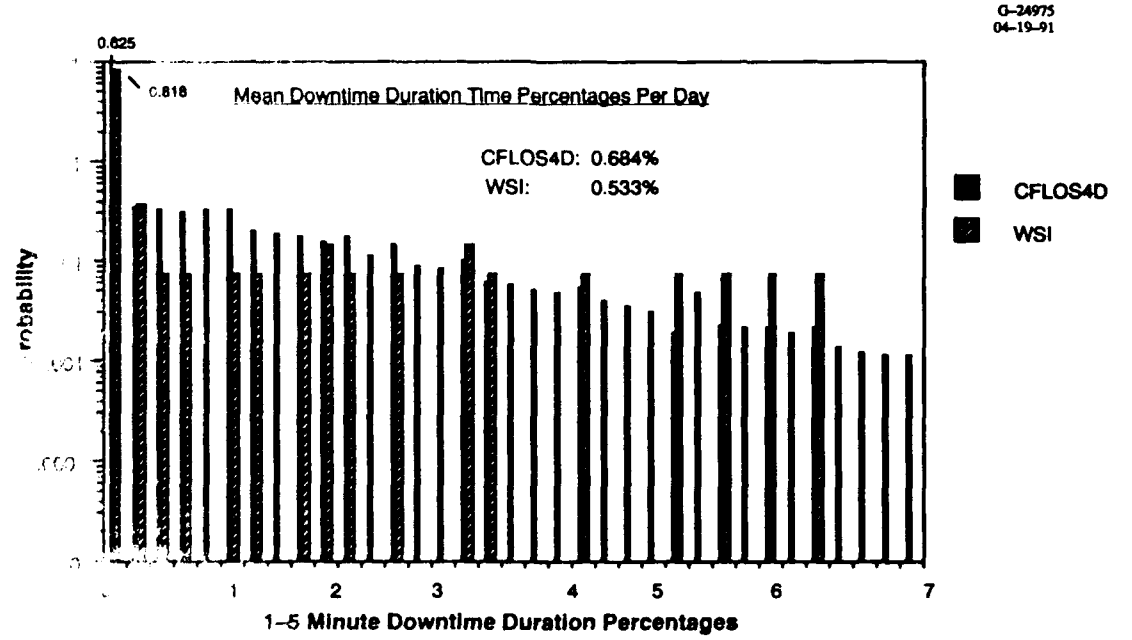


Figure 3.4-5b Case II Histogram for 1-5 Minute Category Percentages

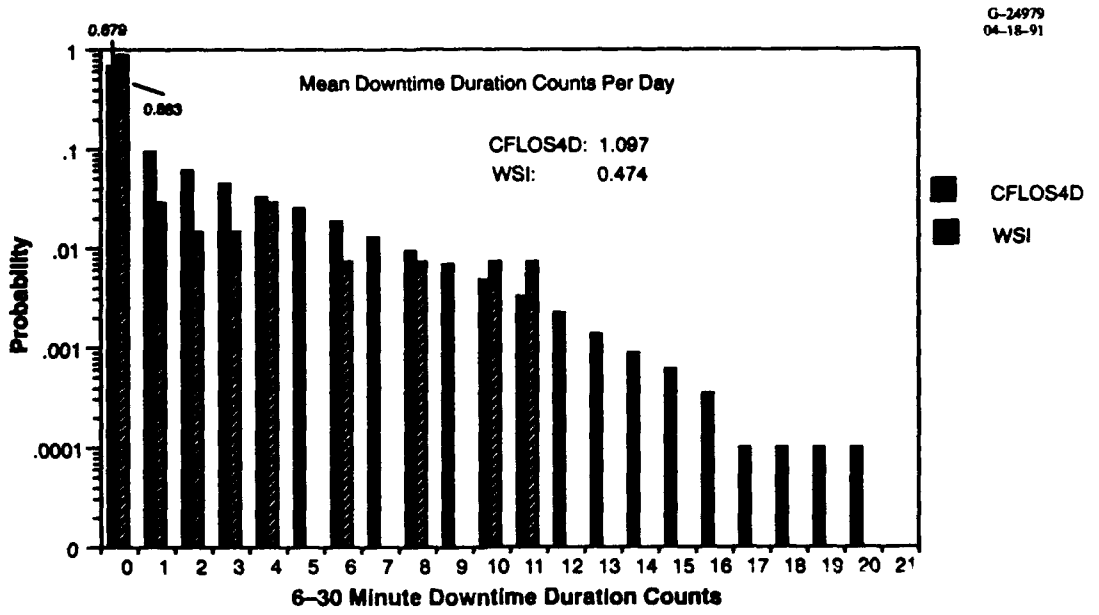


Figure 3.4-6a Case II Histogram for 6-30 Minute Category Counts

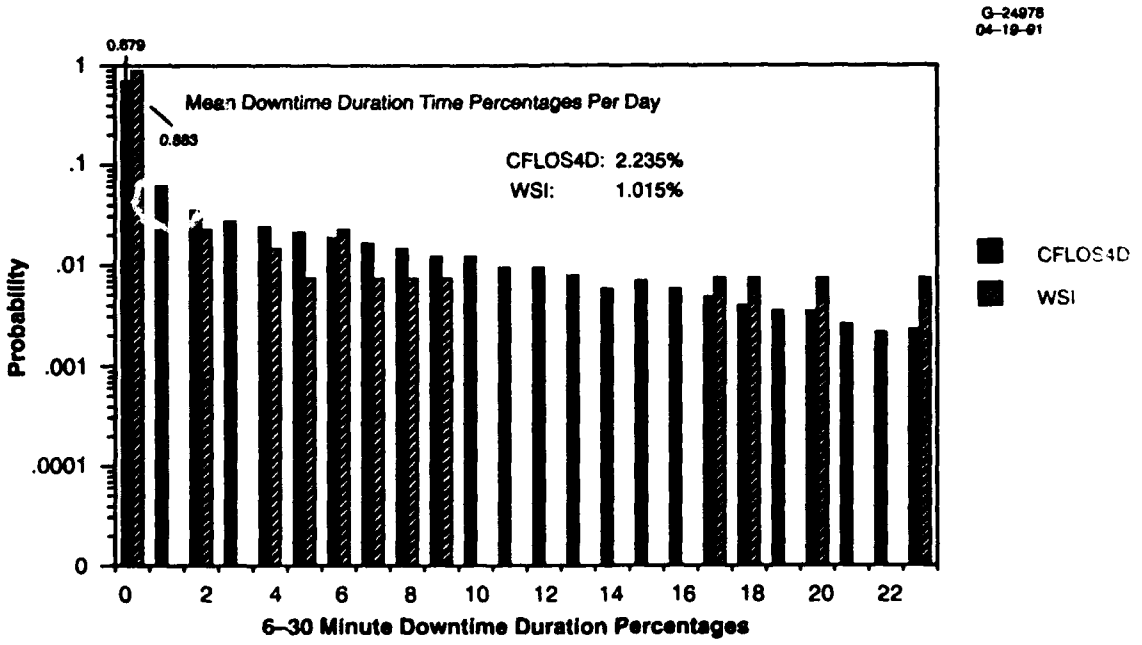


Figure 3.4-6b Case II Histogram for 6-30 Minute Category Percentages

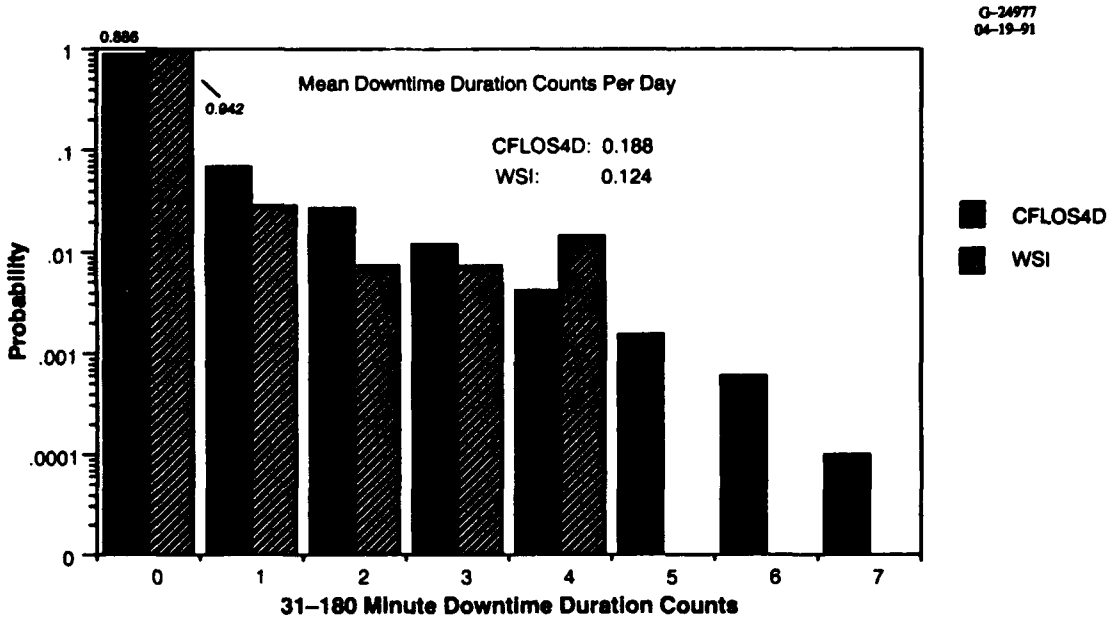


Figure 3.4-7a Case II Histogram for 31-180 Minute Category Counts

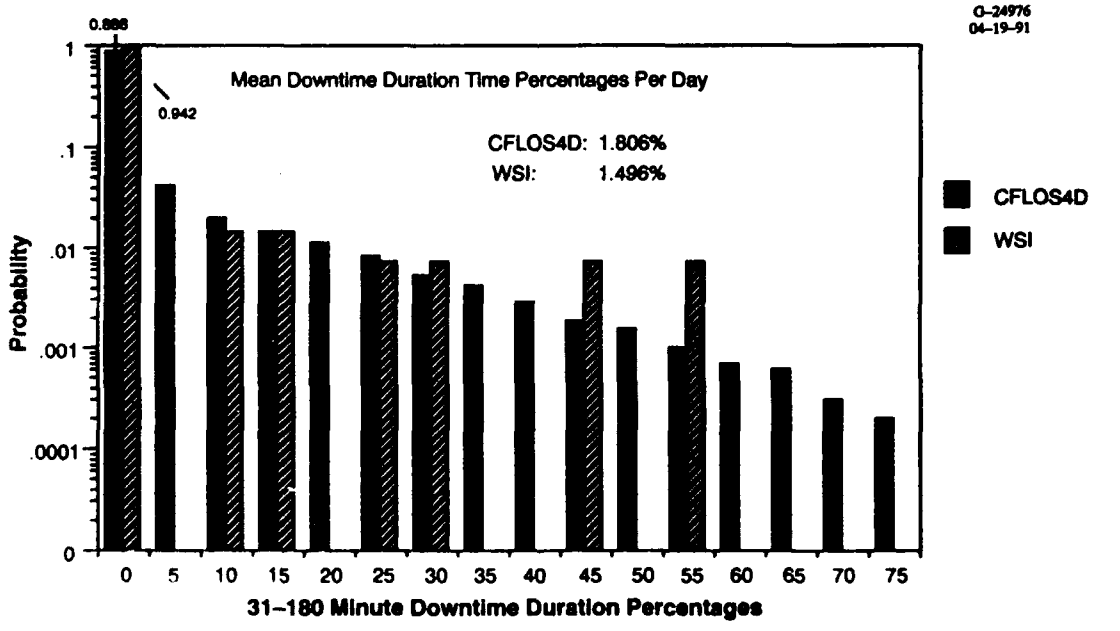


Figure 3.4-7b Case II Histogram for 31-180 Minute Category Percentages

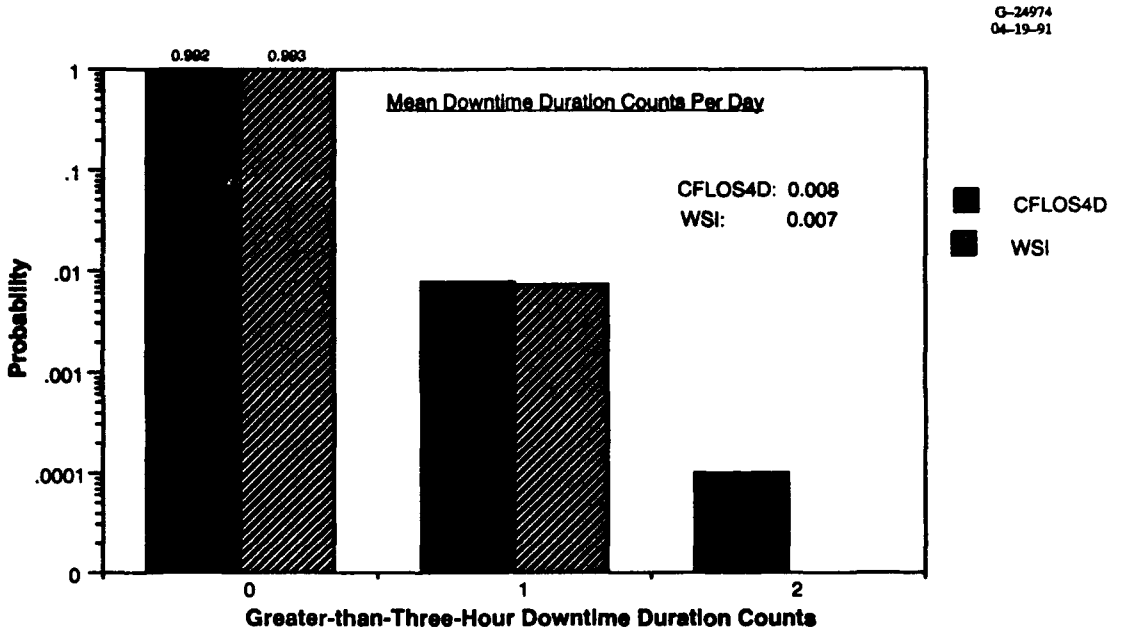


Figure 3.4-8a Case II Histogram for Greater-than-Three-Hour Category Counts

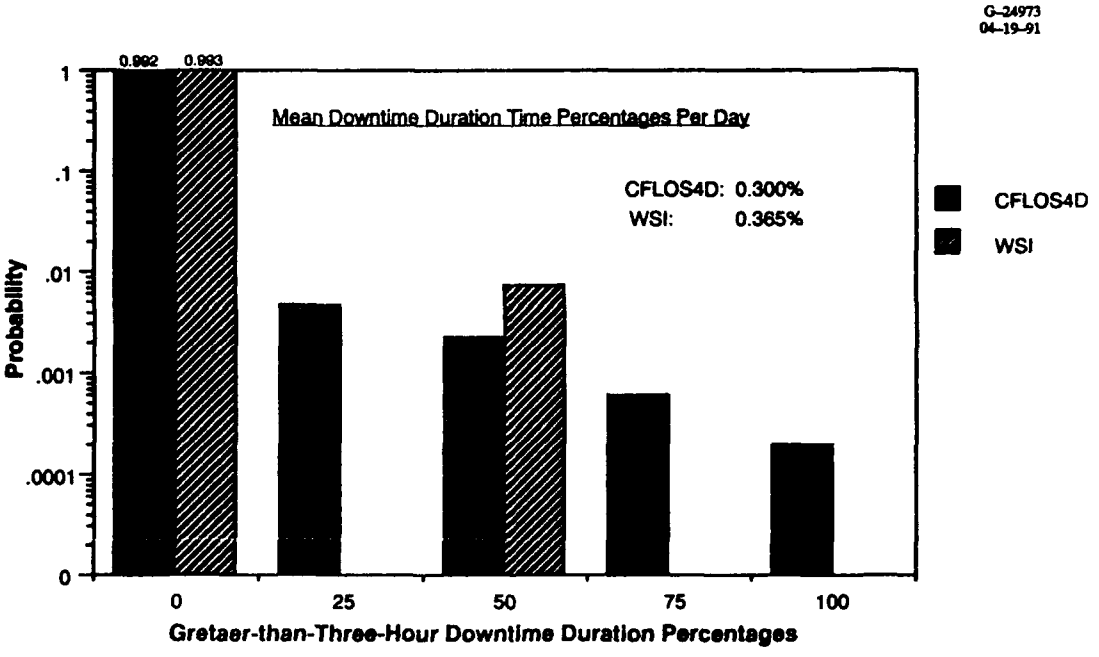


Figure 3.4-8b Case II Histogram for Greater-than-Three-Hour Category Percentages

Table 3.4-1 **Estimated Data/Model Mean Downtime Count Ratios ($\hat{\psi}$)**

DOWNTIME CATEGORY (IN MINUTES)	CASE I	CASE II	CASE III	CASE IV
1-5	1.55	1.01	1.60	1.28
6-30	0.48	0.43	0.50	0.48
31-180	0.58	0.66	0.60	0.64
>180	0.52	0.88	0.65	0.80

indication as to the cause of acceptance or rejection of H_0 via formal KS goodness-of-fit testing. The values shown clearly suggest that CFLOS4D underestimates short downtime duration frequencies and overestimates longer downtime duration frequencies.

The results of the KS goodness-of-fit test for the consistency of CFLOS4D histograms and WSI histograms are presented in terms of *test tail probabilities*. The relationship between a test tail probability, p_{TT}, and the specified false rejection probability (α) is as follows. In testing at a significance level α, rejecting the null hypothesis H_0 if the observed KS statistic D is greater than the threshold t is equivalent to rejecting H_0 if the test tail probability p_{TT} is greater than α. Statisticians often report the tail probability p_{TT} associated with the outcome of a particular test rather than just α and the reject or accept decision. This is because the size of the tail probability carries more information about the strength of evidence for or against H_0 and because the selection of α is quite subjective.

Figure 3.4-9 presents test tail probabilities of the KS statistic for Case I downtime duration counts, for all four downtime categories. (A similar plot is obtained when time percentages are used instead of counts.) An interpretation of the β probabilities indicated adjacent to the curves will be given shortly. The test tail probabilities in Fig. 3.4-9 were computed via Monte Carlo simulation in a procedure inverse to the computation of test thresholds outlined in Fig. 3.3-1. Each of the four curves in Fig. 3.4-9 (corresponding to the four downtime duration categories) is defined over a range of tolerance ratios ψ. (Recall that a tolerance ratio $\psi = 1.1$, for instance, implies an allowable deviation of 10% between model and data-derived downtime count means in accepting the model/data consistency hypothesis.) For example, the probability that the Case I greater-than-three-hour count KS statistic would be greater than or

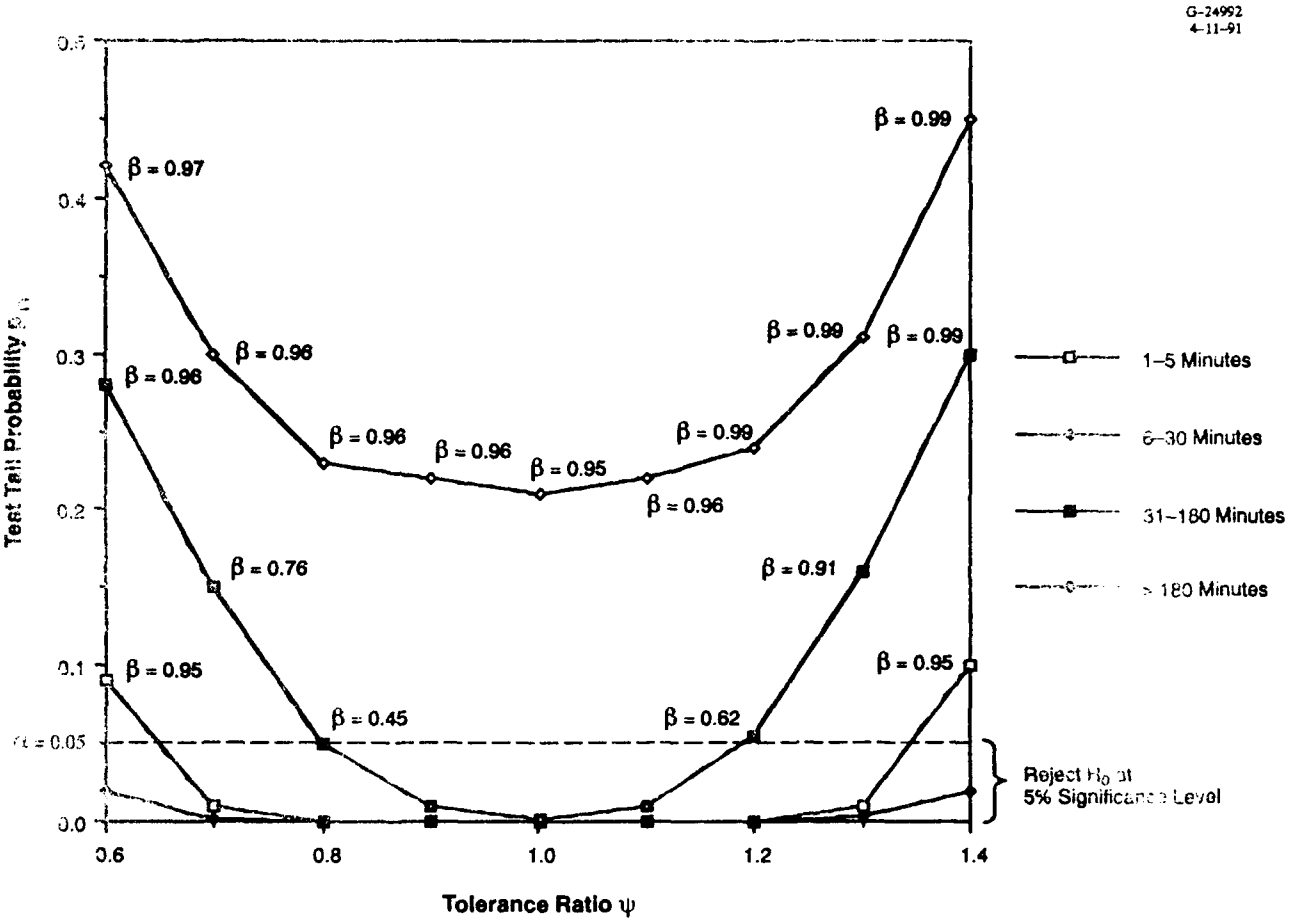


Figure 3.4-9 Case I Test Tail Probabilities for Various Levels of Tolerance

equal to its computed value, *given* that H_0 is true and $\psi = 1.1$, is 0.22. That is, *if* the CFLOS4D and true (but unknown) WSI downtime count histograms are within a prescribed mean downtime count deviation [illegible]%, *then* the probability of the computed KS statistic D exceeding its observed value is 0[illegible]. This probability is relatively large and indicates that consistency for the Case I greater-than-three-hour category cannot be rejected (since a smaller test tail probability would imply the observed D value to be less [illegible]. In contrast, the probability that the Case I KS statistic for the 6-30 minute downtime category would coincide with or exceed its computed value, given H_0 is true, is less than 0.03 for even large ψ = 1.4. This can be regarded as extremely strong evidence against consistency of CFLOS4D and WSI count distributions for the 6-30 minute downtime category, which is not surprising since the CFLOS4D and WSI probabilities of zero downtime counts differ so widely.

Table 3.4-2 presents reject (R) or accept (A) decisions as functions of various false rejection probabilities α and tolerance ratios ψ, based on the test tail probabilities indicated in Fig. 3.4-9. One

Table 3.4-2 Case I Reject (R) or Accept (A) Decisions

α	1-5 MINUTES		6-30 MINUTES		31-180 MINUTES		>180 MINUTES	
	ψ = 1.2	ψ = 1.3	ψ = 1.2	ψ = 1.3	ψ = 1.2	ψ = 1.3	ψ = 1.2	ψ = 1.3
0.01	R	A	R	R	A	A	A	A
0.05	R	R	R	R	A	A	A	A
0.10	R	R	R	R	R	A	A	A

sees in the 31-180 minute category, for instance, that model/data consistency is acceptable given an error tolerance of 20% at α = 0.05 though *not* at α = 0.10. From Table 3.4-1, $\hat{\psi}$ = 0.58, i.e., the ratio of means is quite small in this case, consistent with the fact that p_{TT} is also quite large. Keep in mind, however, that the test procedure is based on the KS statistic, i.e., the estimated mean ratio does not directly enter the statistical decision whether to accept or reject. The parameter ψ, as stated in Section 3.2, is useful in prescribing desired test tolerance or in characterizing the performance of the test, but its estimated value is not an explicit factor in the KS criterion. The KS criterion was used because of its sensitivity to a wide range of deviations between model and data distributions (Refs. 11 and 12).

One could, of course, specify false rejection probability α = 0.0 and accept model/data consistency always (with statistical confidence equal to 1.0). The statistical *power* inherent in such an approach, however, would be equal to 0.0; i.e., the false acceptance probability β = 1.0 for even large discrepancies between model and data distributions. Proper design of a statistical test requires one to balance the relative sizes of α and β, subject to the available WSI sample size N and to the specified tolerance ratio ψ. Given the limited nature of the present WSI data set, the false acceptance probabilities β are, in fact, very large for even reasonable false rejection probabilities α = 0.05 and 20% error tolerance (Fig. 3.4-9). For example, the probability of accepting the hypothesis that the WSI and CFLOS4D downtime count histograms in Fig. 3.4-3a (31-180 minutes, Case I) are consistent when, in truth, they are *inconsistent*, exceeds 0.40. The most natural solution to this shortage of statistical power is to seek an extended WSI data set.

3.4.2 Data Requirements

We now shift focus from quantifying decision error probabilities for the existing WSI data set and ask instead: What WSI sample size N *would be* needed to reduce appropriately the probability, β, of

failing to detect a mean deviation exceeding ψ ? Figures 3.4-10 to 3.4-13 provide an answer, based on the CFLOS4D and WSI histograms for Case I only, for each of the four downtime duration categories. An impression of the sensitivity of N to varying α and β is given. Since there is a trade-off between the decision error probabilities (i.e., for fixed N, α decreases if and only if β increases at a fixed ψ), it follows that

- increasing both α and β yields a smaller required N
- decreasing both α and β yields a larger required N
- increasing α (β) and keeping β (α) constant yields a smaller required N
- decreasing α (β) and keeping β (α) constant yields a larger required N.

Figures 3.4-10 to 3.4-13 serve to make the above intuitive observations more precise.

The WSI sample size corresponding to Case I at present is 242 days; therefore, with α = 0.1 and β = 0.25, one can expect to detect a deviation between CFLOS4D and WSI means of no less than 20-30% for the 1-5, 6-30 and 31-180 minute downtime duration categories. With one additional year of

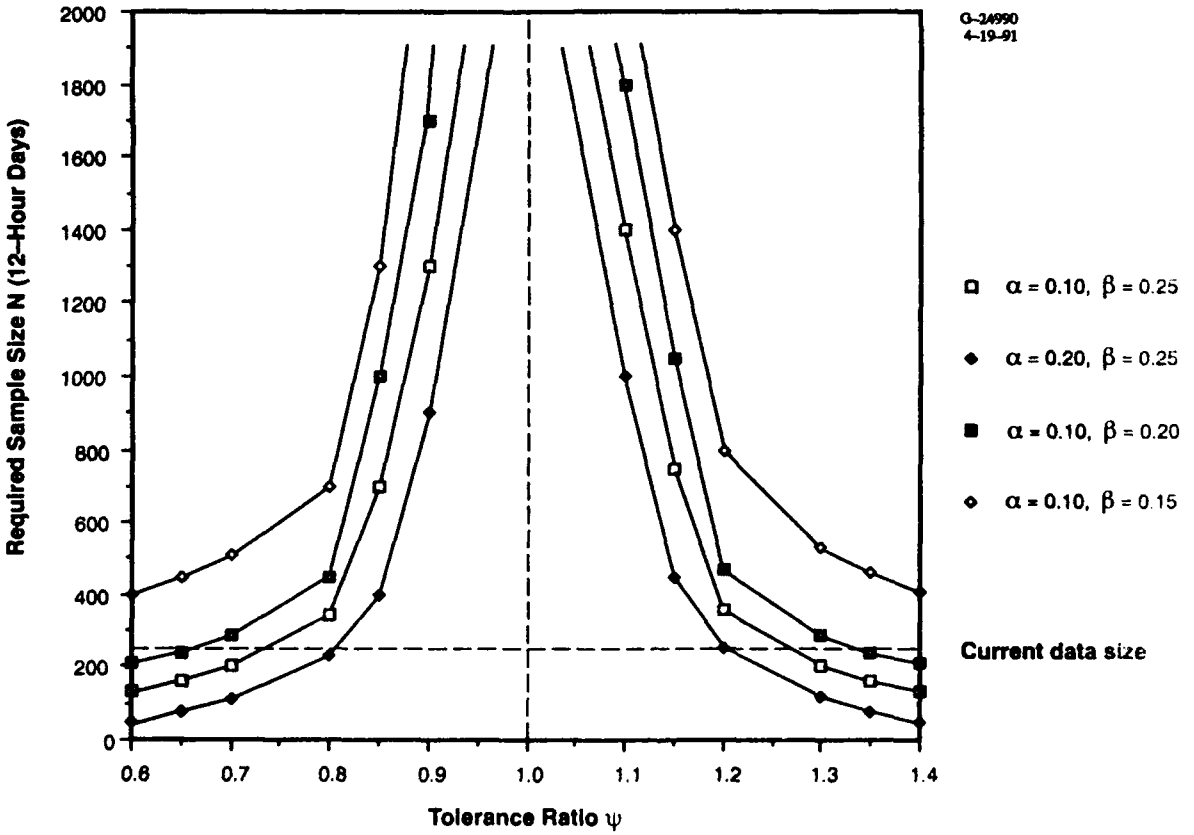


Figure 3.4-10 Data Requirements Analysis: Case I 1-5 Minutes Downtime Category

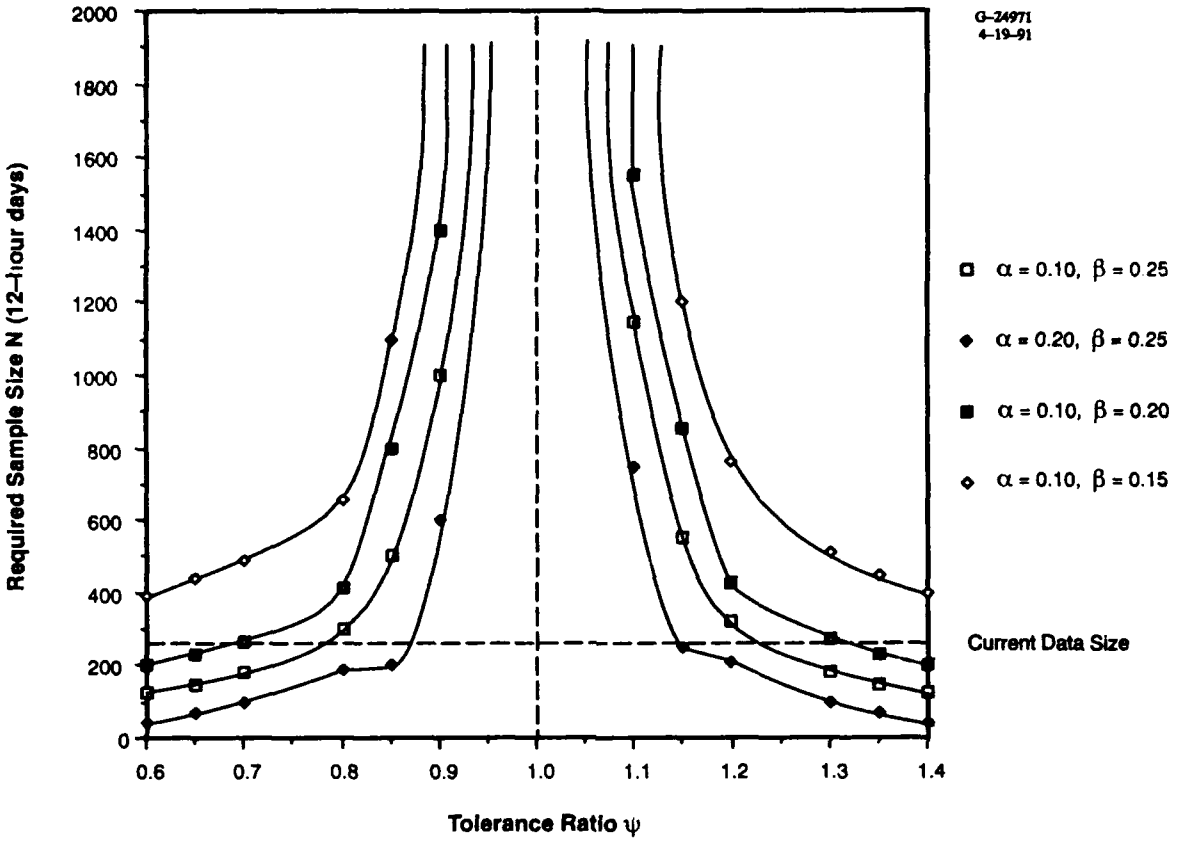


Figure 3.4-11 Data Requirements Analysis: Case I 6-30 Minutes Downtime Category

WSI data (365 days), mean deviations of no less than 15-20% can be detected; with two additional years of WSI data, mean deviations of no less than 10-15% can be detected. (In fact, from Table 3.4-1, the mean deviations are all estimated to be 55-60% and consistency is, indeed, rejected for all three categories with $\alpha = 0.10, \psi = 1.2$ (Fig. 3.4-9).)

The greater-than-three-hour downtime duration category requires special treatment. On one hand, the difference between the CFLOS4D zero downtime probability of 0.9773 and the WSI zero downtime probability of 0.9876 is only 0.0103, which seems very small. On the other hand, the mean deviation is estimated to be 48%, which seems very large (*twice* as many downtimes exceeding three hours, on average, are predicted by the model over that observed). Consistency is accepted (Fig. 3.4-9, with $\alpha = 0.10, \psi = 1.2$) for the present data set; a minimum of two years of additional data would be needed to reliably detect mean differences of 40-50% in this category and even then false acceptance (β) probabilities would be approximately 0.5.

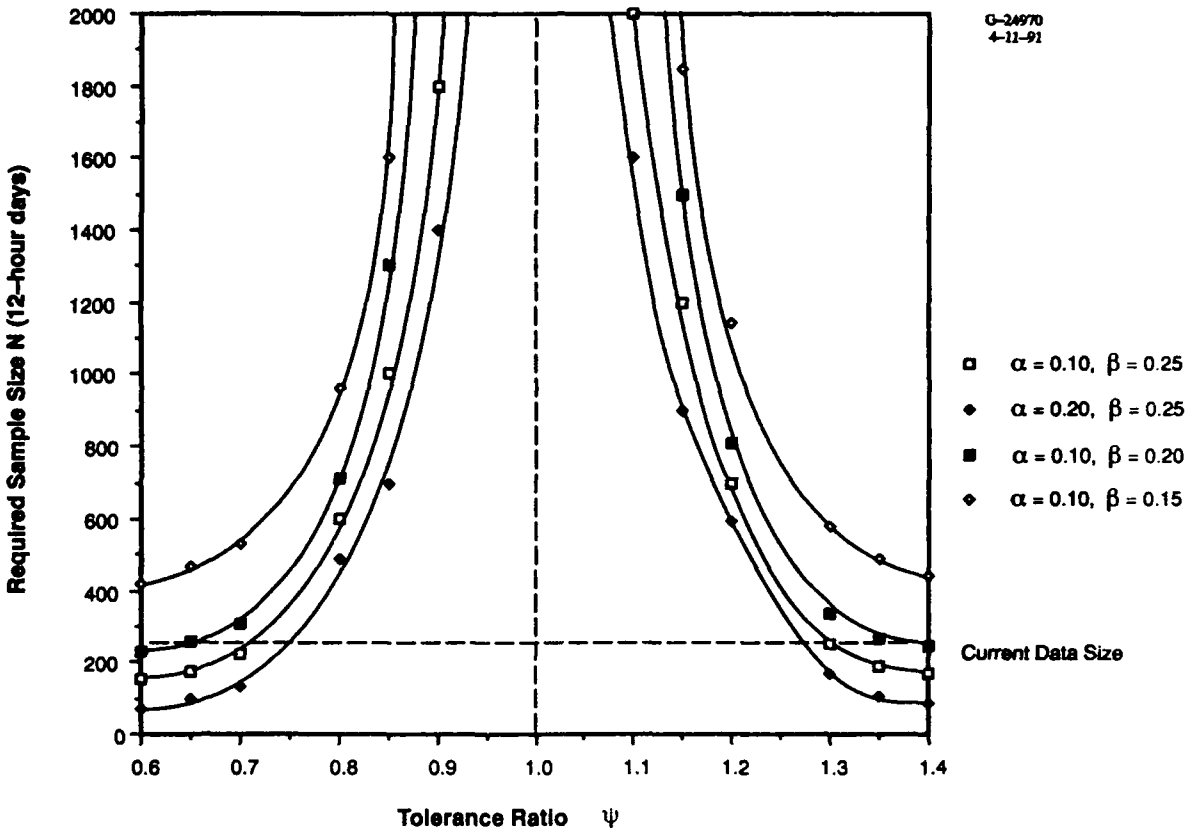


Figure 3.4-12 Data Requirements Analysis: Case I 31-180 Minutes Downtime Category

Results in these figures were computed by a data requirements algorithm (Fig. 3.4-14). The algorithm input is a desired false acceptance probability, β_{des}. The algorithm computes iteratively the minimum required sample size N, for specified α, that distinguishes distributions of separation with probability approximately equal to β_{des}. During the course of the iterations, if a resulting β estimate is *larger* than β_{des}, then N is too small and should be increased. If the β estimate is *smaller* than β_{des}, then N is unnecessarily large and should be decreased. The β probability algorithm is called again based on a new candidate sample size N and repeated until β estimates agree suitably with β_{des}.

A consequence of these data requirements is the need to more fully exploit the entire grid of WSI clear/cloudy data at each minute. This may be accomplished by using several pixels per (multi-site) image. The *effective* increase in sample size by incorporating this procedure is a function of inter-pixel spatial dependence; effective gains are small when correlation is large and vice versa. For instance, if three pixels can be found with spatial correlation coefficient r between each pair of pixels, then the

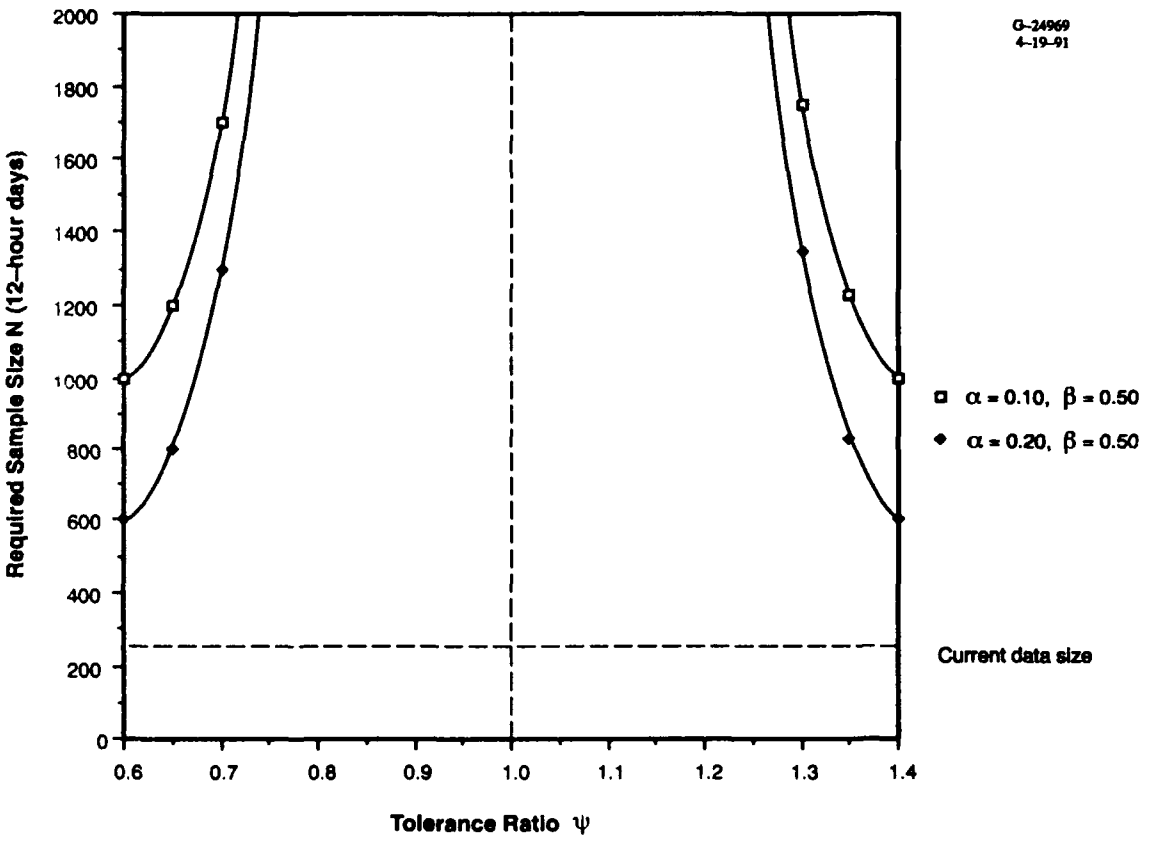


Figure 3.4-13 Data Requirements Analysis: Case I Greater-than-Three-Hours Downtime Category

effective increase in sample size is (2 - 2 r) / (1 + 2 r). Thus, if r = 0.25, the sample size effectively doubles; if instead r = 0.50, the effective increase in sample size is only 50%. Similar formulas for effective sample size as functions of spatial correlation, for any number and arrangement of WSI pixels, have been determined and could be employed in future CFLOS4D/CFARC validation activities (Ref. 7).

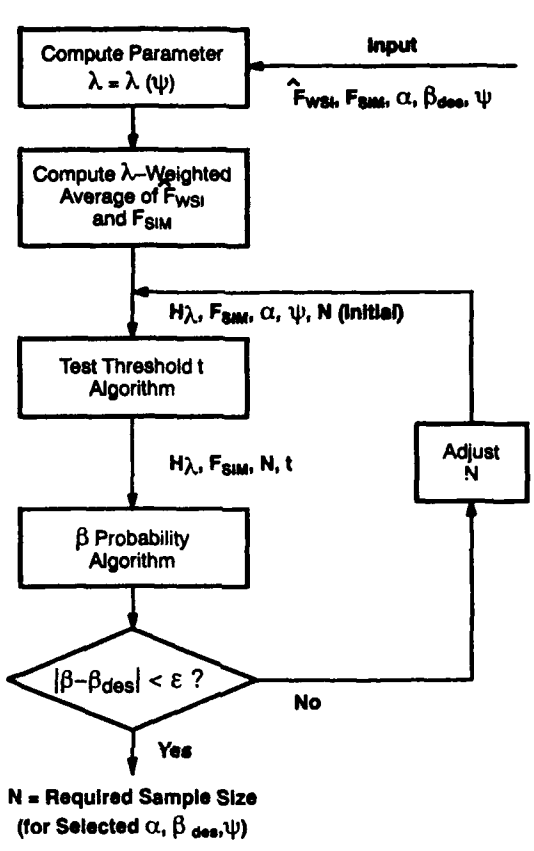


Figure 3.4-14 Algorithm for Data Requirements

4. MODEL COMPONENT INVESTIGATIONS

In this section individual model components comprising the CFLOS simulators are evaluated with respect to corresponding characteristics derived from the WSI data set. These evaluations provide insight into the accuracy, relative to the WSI data set, of models which may be used individually for other applications. Three fundamental model elements are addressed:

- probability of a cloud-free line-of-sight (PCFLOS)
- fractional sky cover spatial correlation
- fractional sky cover temporal correlation

Time did not permit thorough investigations of sky dome CFLOS spatial and temporal behavior. However, related to investigations of CFLOS temporal characteristics, persistence and recurrence CFLOS frequencies determined from the data set are compared with the appropriate models. Downtime duration distributions for a single site are also investigated. Detailed analysis of CFLOS temporal and spatial correlation must await future studies.

4.1 PROBABILITY OF A CLOUD-FREE LINE-OF-SIGHT

The PCFLOS model incorporated into the CFLOS simulators is derived from that of Allen and Malick (Ref. 13). It is commonly termed the SRI model. The model provides an analytic expression for PCFLOS as a function of sky cover and zenith angle and is based on the work of Lund and Shanklin (Ref. 14). The model is given by

$$\mathrm{PCFLOS}(s) = (1 - s(1 + 3s)/4)^{(1+(0.55-s/2)\tan(\beta))} \qquad (4.1\text{-}1)$$

where s is fractional sky cover in tenths and β is the zenith angle.

Data-derived probabilities of CFLOS were computed directly from the WSI data at a single site and for specific sky dome points (corresponding to known zenith angles) by adding the number of clear occurrences at a point and dividing by the total number of clear and cloudy occurrences. Only total cloud data were used in the investigation. As each sky dome point is associated with a zenith angle and each sky dome image is associated with a computed sky cover value, the calculations were organized in zenith angle and sky cover category (tenths) bins to enable easy comparison with the model.

Figure 4.1-1 depicts the variation of data-derived PCFLOS *at zenith* with sky cover as determined from the WSI Columbia data set. Also shown are the SRI model values and the complement of sky cover, or cloud-free fraction, which is a first-order model for PCFLOS. As indicated in the figure, the sample size exceeded 150,000 data points. (Of course, the data are highly correlated). The most interesting feature of this comparison is the unexpected dip of the data-derived PCFLOS curve below the cloud-free fraction. If real, this effect could be indicative of predominant frontal weather patterns moving across the region such that zenith is cloud-covered when the cloud fraction is high. The possibility of defective data, however, cannot be ignored. Although not shown, a similar comparison was observed using WSI data from the Kirtland site.

Employing the same Columbia data set, Fig. 4.1-2 provides another data/model comparison for PCFLOS. The SRI model at three different sky cover fractions is compared with data-derived PCFLOS vs. elevation angle. Consistent with the results of Fig. 4.1-1, the data-derived curve dips well below the model curve with increasing elevation angle at high sky cover fraction (0.8). Higher data-derived PCFLOS values are seen at middle elevation angles and lower data-derived PCFLOS values are seen at the lower elevation angles, relative to model values. Note, however, that the data quality problems known to exist near the horizon may make the comparison at low elevation angles meaningless.

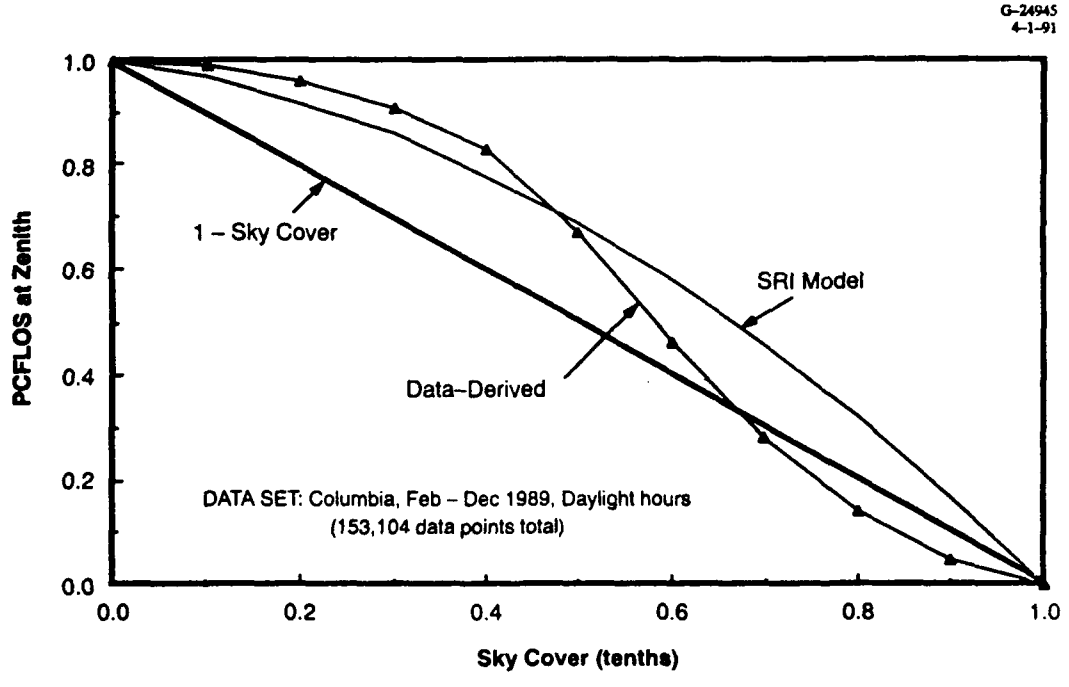


Figure 4.1-1 Comparison of WSI Data-Derived and SRI Model PCFLOS at Zenith

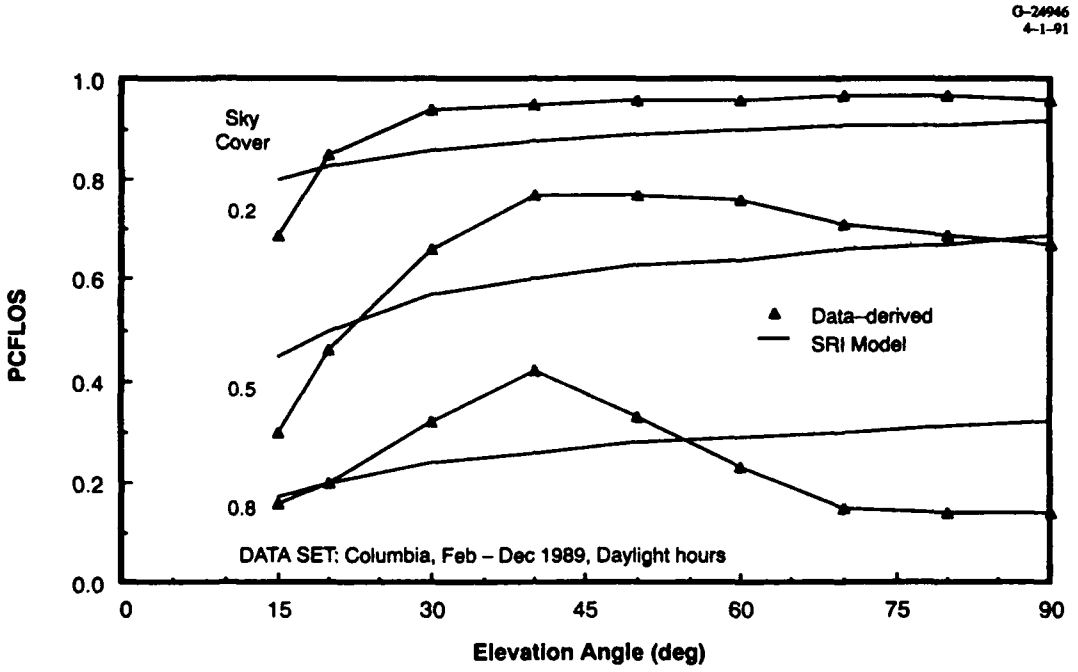


Figure 4.1-2 PCFLOS vs. Elevation Angle, WSI Data and SRI Model

4.2 SKY COVER SPATIAL CORRELATION

In developing the spatial correlation model for use in the CFLOS simulators (Ref. 3) the spatial correlation of sky cover as determined from thousands of U.S. surface observations of sky cover played a central role. In fact, the CFLOS spatial correlation model is comprised of a weighted sum of small and large scale components, and the sky cover spatial (site-to-site) correlation is the large scale component. The sky cover spatial correlation model is analytically expressed as

$$\varrho = 1 - \frac{8}{\pi}\left(\frac{d}{\Lambda}\right) + \frac{3}{2}\left(\frac{d}{\Lambda}\right)^2 \quad \text{for } d \le \Lambda \tag{4.2-1}$$

where d is the distance in kilometers between the sites and Λ is the large scale wavelength fixed in the CFLOS simulators at 1440 km.

The simulator models employ *tetrachoric* correlation (Ref. 3), which is computed dichotomously from sky cover: the sky is identified as clear if sky cover is less than or equal to 50% and identified as cloudy otherwise. The tetrachoric correlation computation implicitly assumes that the underlying variable (sky cover) is normally distributed. The procedure tabulates, in a two-by-two contingency table, the

occurrences of the sky cover pairs clear-clear (00), clear-cloudy (01), cloudy-clear (10), and cloudy-cloudy (11), from the two sites under consideration. Given the completed contingency table with elements c_{00}, c_{01},c_{10}, and c_{11}, tetrachoric correlation is computed by

$$\varrho_{TET} = \sin\left[\frac{\pi}{2}\,\frac{\sqrt{c_{00}\,c_{11}} - \sqrt{c_{01}\,c_{10}}}{\sqrt{c_{00}\,c_{11}} + \sqrt{c_{01}\,c_{10}}}\right] \tag{4.2-2}$$

If the underlying continuous-valued random variables (sky cover fraction in this case) were actually normally distributed, then Eq.4.2-2 is an excellent estimator of the correlation coefficient between bivariate normal random variables. Since sky cover fraction is not normally distributed (the distribution of sky cover is usually U, L or J - shaped (Ref. 1)), use of Eq. 4.2-2 should be viewed with caution.

From the fractional sky cover values available from each WSI image, however, the more conventional Pearson product-moment correlation was also computed. The site-to-site sky cover product-moment correlation is given by

$$\varrho_{PEARS} = \frac{\sum_{i=1}^{n} (sc_{i1} - \mu_1)\,(sc_{i2} - \mu_2)}{\sqrt{\sum_{i=1}^{n} (sc_{i1} - \mu_1)^2}\;\sqrt{\sum_{i=1}^{n} (sc_{i2} - \mu_2)^2}} \tag{4.2-3}$$

where n is the number of sample pairs, sc_{ij} is the sky cover at time i and site j, $1 \le i \le n$, $1 \le j \le 2$, and where

$$\mu_j = \frac{1}{n} \sum_{i=1}^{n} sc_{ij} \tag{4.2-4}$$

denotes the mean sky cover at site j.

Site-to-site sky cover spatial correlations are listed in Table 4.2-1 for Kirtland and White Sands C-Station. This is the only site pair, for which we have data, for which we observe non-zero correlations. Shown in Table 4.2-1 are tetrachoric correlations derived from the model, the WSI data, and surface observations concurrent with the WSI data. The WSI data-derived product-moment correlation with associated 90% confidence limits, computed using the Fisher z-transformation (Ref. 15), is also indicated. The sample size exceeded 115,000 1-minute data pairs covering 230 days of daylight hours from March through November, 1989. High temporal correlations inherent in the data (see Section 4.3) were

Table 4.2-1 Sky Cover Spatial Correlations

Kirtland — White Sands (C-Station)

Tetrachoric Correlation	
CFLOS4D Model	0.58
WSI Data-derived	0.52
Surface Observations	0.57
WSI Product-Moment Correlation	0.42 (0.29, 0.53)*

*90% Confidence Limits

necessarily acknowledged in computing the 90% confidence limits of the product-moment correlation estimates.

Although the tetrachoric correlations match up fairly well, the product-moment correlation, traditionally a better estimator of correlation between the random quantities of interest, falls significantly below the tetrachoric correlation values. This suggests that the spatial correlation as modeled in the CFLOS simulators is too high.

4.3 SKY COVER TEMPORAL CORRELATION

Similar to the structure of the CFLOS spatial correlation model, the CFLOS temporal correlation model consists of a weighted sum of short and long scale components. The Lund data set (Ref. 14) was a key resource in assembling the CFLOS temporal correlation model used in the simulators. However, guidance on the long scale component was provided by determining the temporal tetrachoric correlation of sky cover using hourly surface observations from many U.S. sites (Ref. 3). The model is represented by an exponential with a decay constant of 13 hours:

$$\varrho(\delta t) = \exp\left[\frac{-\delta t}{T_D}\right] \tag{4.3-1}$$

where T_D = 13 hours.

Both tetrachoric and product-moment temporal correlation were computed from the WSI sky cover data for comparison with the model. Using the Fisher z-transformation, 90% confidence limits for

the product-moment correlation estimates were determined as well. Also, temporal correlations from hourly sky cover observations were computed from concurrent surface data. The results, plotted vs. time lag in minutes, are shown in Fig. 4.3-1 for the Kirtland site and in Fig. 4.3-2 for the Columbia site.

As was the case for sky cover spatial correlation, the model values match well with the data-derived tetrachoric correlation estimates, but the product-moment correlation estimates suggest that the

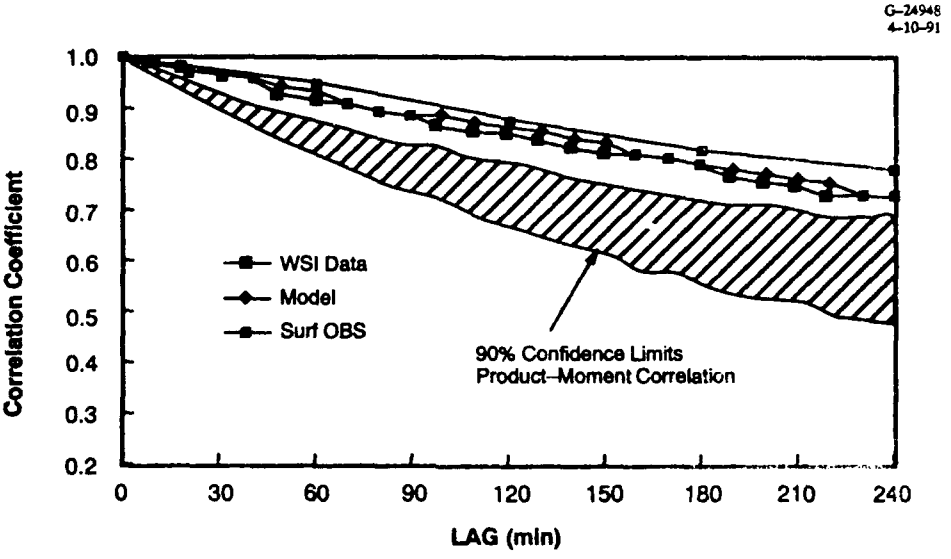


Figure 4.3-1 Sky Cover Temporal Correlation Estimates — Kirtland

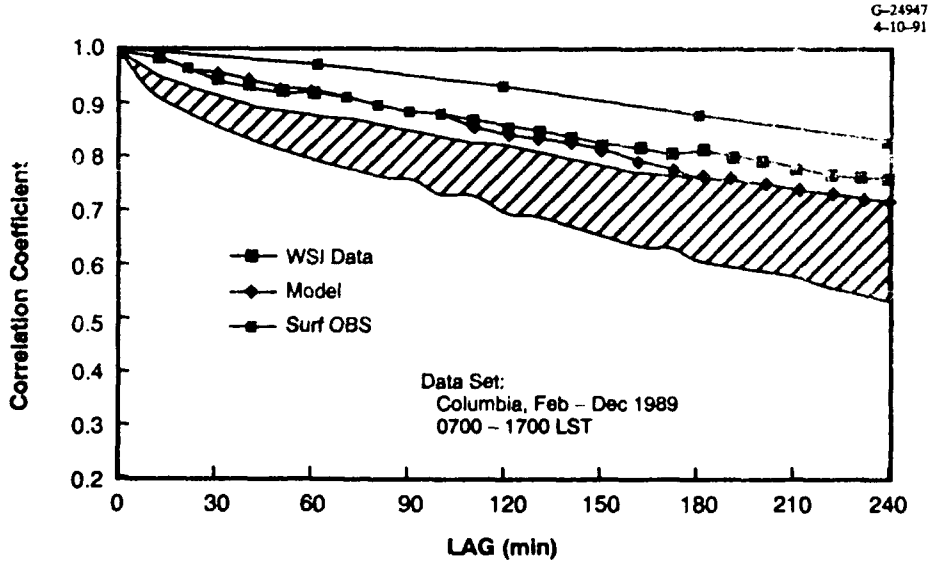


Figure 4.3-2 Sky Cover Temporal Correlation Estimates — Columbia

model values are too high. These estimates, computed over an eleven month time period for which data are available, represent aggregate quantities. Diurnal and seasonal variations have yet to be ascertained.

For the Columbia case (Fig. 4.3-2), the discrepancies between the tetrachoric correlation estimates derived from the WSI data and the surface observations are noticeable, given that the observations are concurrent with the WSI data. Recall from Section 2, however, that surface sky cover observations are coarsely categorized and some inconsistencies between the data sets are known to exist.

4.4 SINGLE-SITE DOWNTIME DURATION DISTRIBUTIONS

For a specified line-of-sight (defined by azimuth and elevation angles) at a given site, a distribution of downtime durations is equivalent to a distribution of cloudy intervals of time. In this section, a determination of downtime duration counts corresponds to system level output for a single site, except that durations of any integral number of minutes, not just for specified duration categories, are tabulated. In this sense results in this section pertain to the temporal character of CFLOS and complement the results of Section 4.3. Also, further insight is gained into the tendency of the simulator to underpredict downtimes of short duration yet overpredict downtimes of longer duration, as reported in Section 3.

Figure 4.4-1 provides a comparison of WSI and CFLOS4D empirical frequency distributions of downtime duration, in minutes, for the Columbia site. Corresponding cumulative distributions are compared in Fig. 4.4-2. Results are shown for durations up to thirty minutes in length; results for longer durations were computed but add little to the model/data comparison. WSI data during daylight hours for the months February through December were used in constructing the WSI distributions, and 1500 CFLOS4D realizations over the WSI data period enabled a reliable determination of the simulator distributions. As discussed in Section 2, cloud/no-cloud data reduction was optimized for the northern sky, away from the sun occultor. Thus, to reduce the influence of possible data errors, the evaluation has been carried out for a particular point in the northern sky at an elevation angle of 60 degrees.

Consistent with the findings of Section 3, the data clearly indicate higher occurrences of shorter downtime durations, or cloudy lines-of-sight, relative to simulator results. The log scale in Fig. 4.4-1 emphasizes the deviations at the longer durations, but Fig. 4.4-2 clearly reveals the significance of the deviations at the shorter durations. Similar comparisons are provided in Figs. 4.4-3 and 4.4-4 based on the Kirtland data set. These results strengthen the indication of Section 4.3 that the simulator temporal correlation model may be too strong.

To add further credence to the hypothesis that unrealistically high model temporal correlations yield the characteristics seen in Fig. 4.4-1 through 4.4-4, the simulator was exercised over the Columbia

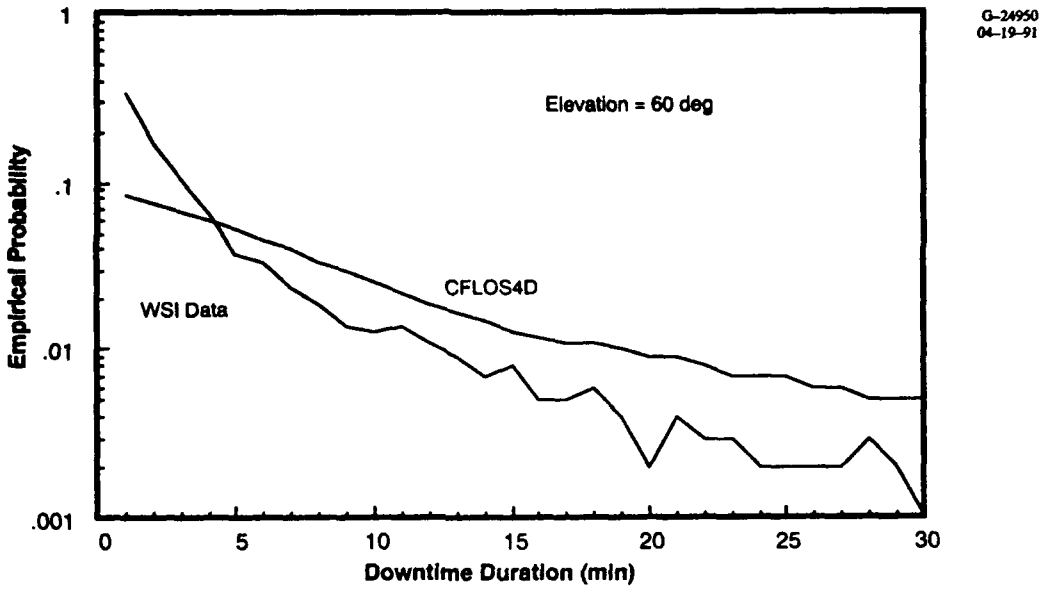


Figure 4.4-1 Distribution of Downtime Durations — Model/Data Comparison (Columbia)

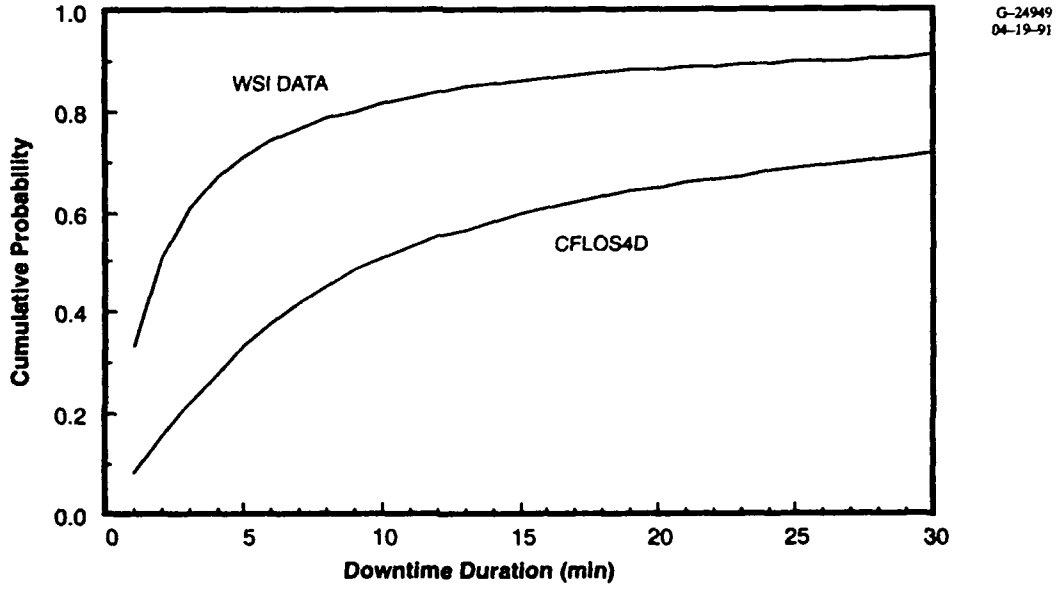


Figure 4.4-2 Cumulative Distribution of Downtime Durations — Model/Data Comparision (Columbia)

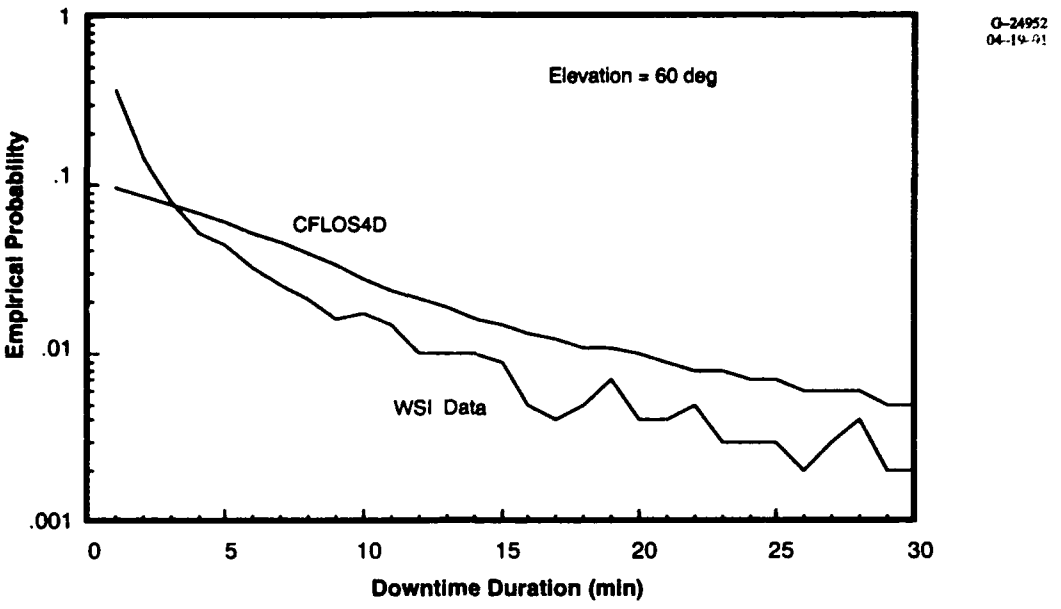


Figure 4.4-3 Distribution of Downtime Durations — Model/Data Comparison (Kirtland)

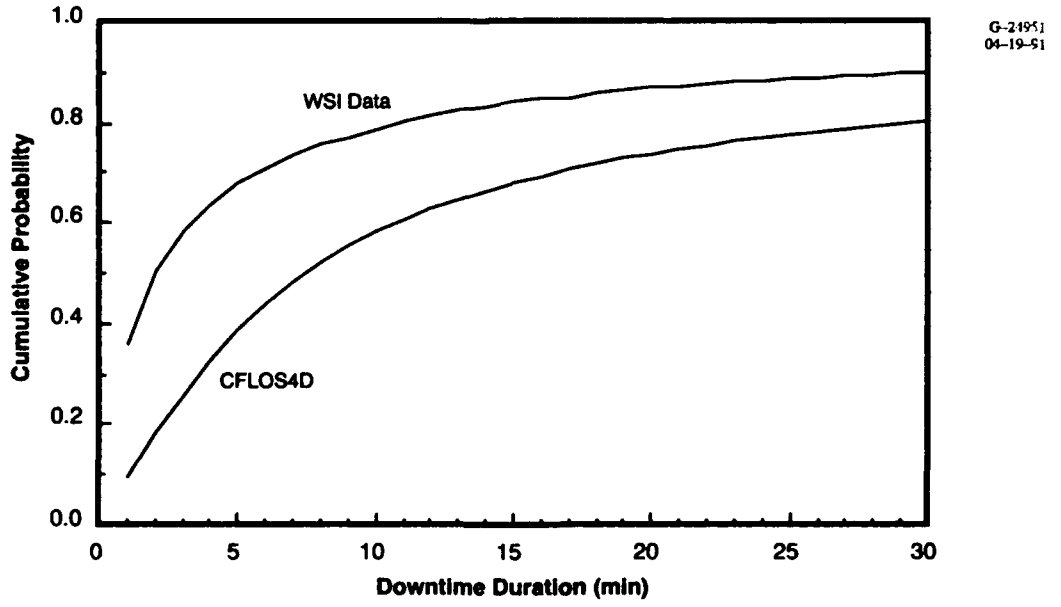


Figure 4.4-4 Cumulative Distribution of Downtime Durations — Model/Data Comparision (Kirtland)

scenario but with an adjusted temporal correlation parameter. This parameter, the short time scale correlation, was adjusted to yield an effective short time scale relaxation time of four minutes. The model value is twenty minutes. As verified in Figs. 4.4-5 and 4.4-6, the downward adjustment in the modeled temporal correlation brings simulator downtime distribution results in line with data characteristics for this single site case. In these figures the data-derived results are denoted by the black triangles to better delineate data and simulator results. Note that this exercise is not intended to determine a proper value of this model parameter, but to show that its adjustment achieves the desired behavior of the simulator.

4.5 CFLOS RECURRENCE AND PERSISTENCE

In support of temporal correlation investigations, CFLOS recurrence and persistence characteristics were computed from the WSI data set. As used here persistence is an uninterrupted sequence of an event (e.g., CFLOS) while recurrence only involves an event recurring at a later time given that the event occurred at the initial time. The main intent is to compare data-derived persistence and recurrence results with the well-known Lund results (Ref. 9). Persistence and recurrence models do not exist in the CFLOS simulators per se, but the Lund data set (about 65 days of whole sky photos at five-minute intervals at Columbia, MO) was used in formulating the simulator CFLOS temporal correlation model (Ref. 3). Indeed, the Lund work has been a critical resource for modeling and assessing cloud impacts on land-based and air-based electro-optical systems in many investigations. One concern, mentioned by Lund in his work, is the effect of the discrete sampling interval (five minutes) on the estimated persistence probability. This can be investigated using the one-minute WSI data.

Figures 4.5-1 and 4.5-2 provide some sample results of CFLOS persistence averaged over 40 points in the northern sky. The empirical persistence probability curves are conditioned on an initial CFLOS and the initial sky cover value indicated in the figure. Total cloud data were used in the evaluation. A significant difference between WSI-derived persistence probabilities computed at one-minute intervals and at five-minute intervals is clearly evident. Differences between the WSI results and those of Lund are also large. Sample variability may be a contributing factor in this difference. Cloud detection errors in both the WSI and Lund data sets may also account for some differences.

Figure 4.5-3 compares empirical recurrence probabilities computed from the WSI data set and the Lund model. Unlike persistence, average clear recurrence probabilities approach the single point probability of a clear line-of-sight (one minus sky cover fraction) as the time span increases. Another observation of importance is the faster fall-off of the WSI-derived result relative to the Lund curve as the

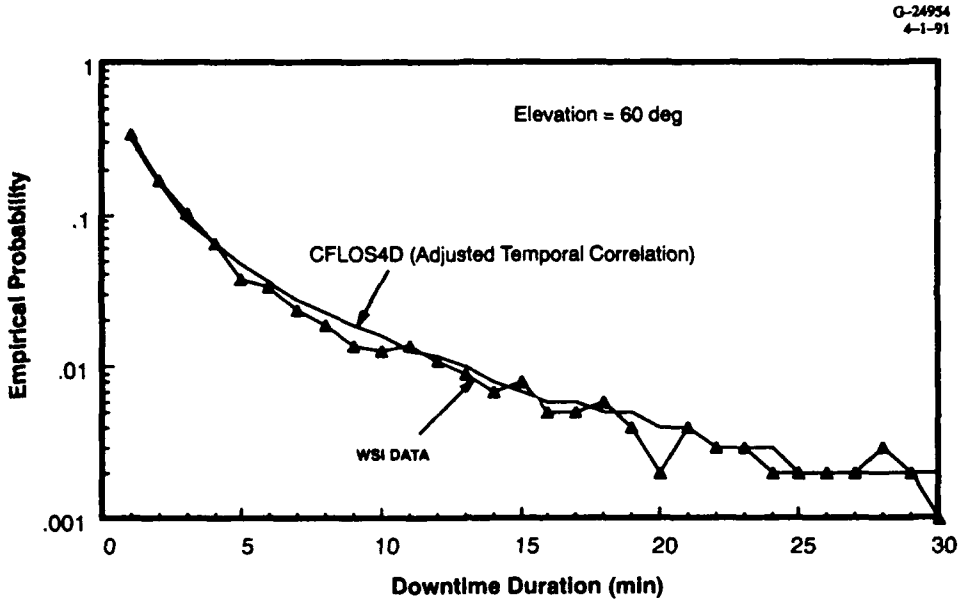


Figure 4.4-5 Distribution of Downtime Durations with Adjusted Model (Columbia)

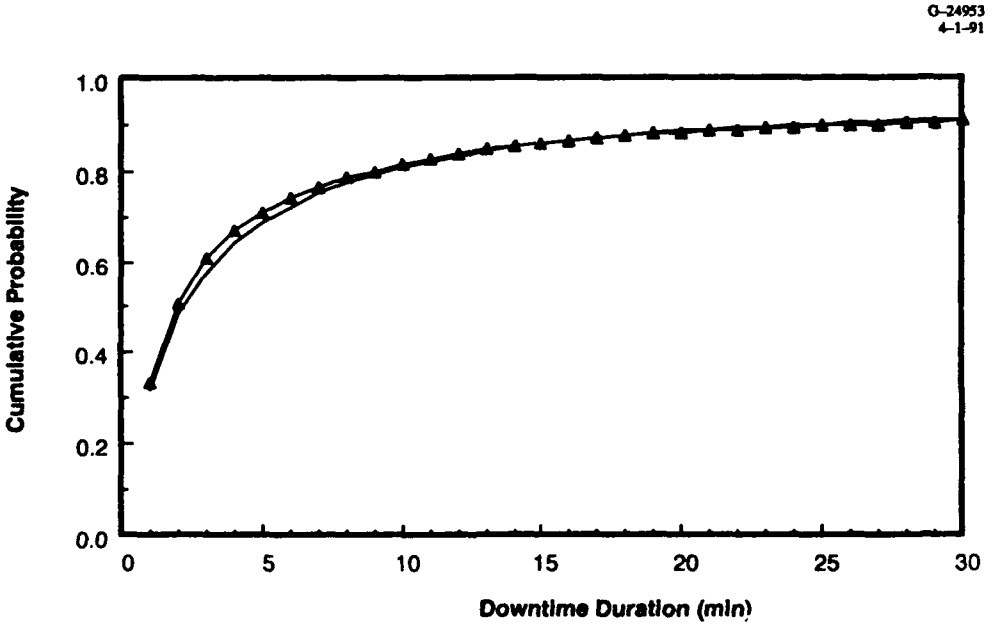


Figure 4.4-6 Cumulative Distribution of Downtime Durations with Adjusted Model (Columbia)

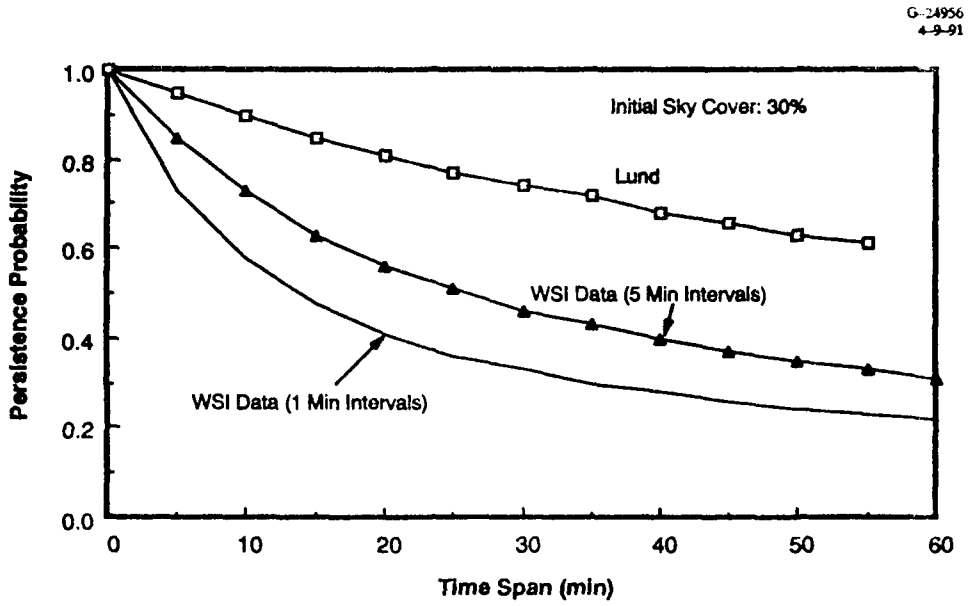


Figure 4.5-1 Persistence Probability — Model/Data Comparison (Columbia)

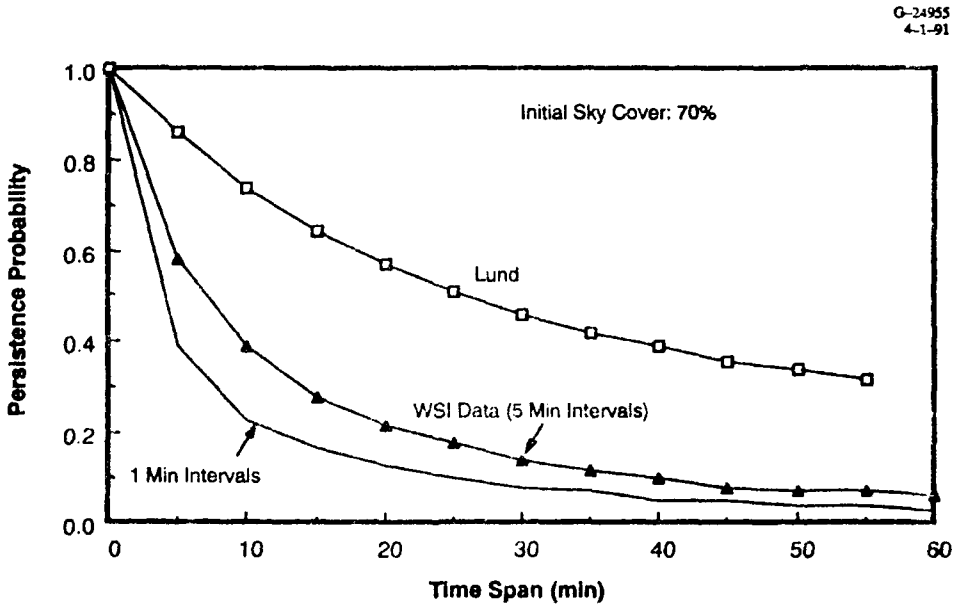


Figure 4.5-2 Persistence Probability — Model/Data Comparison (Columbia)

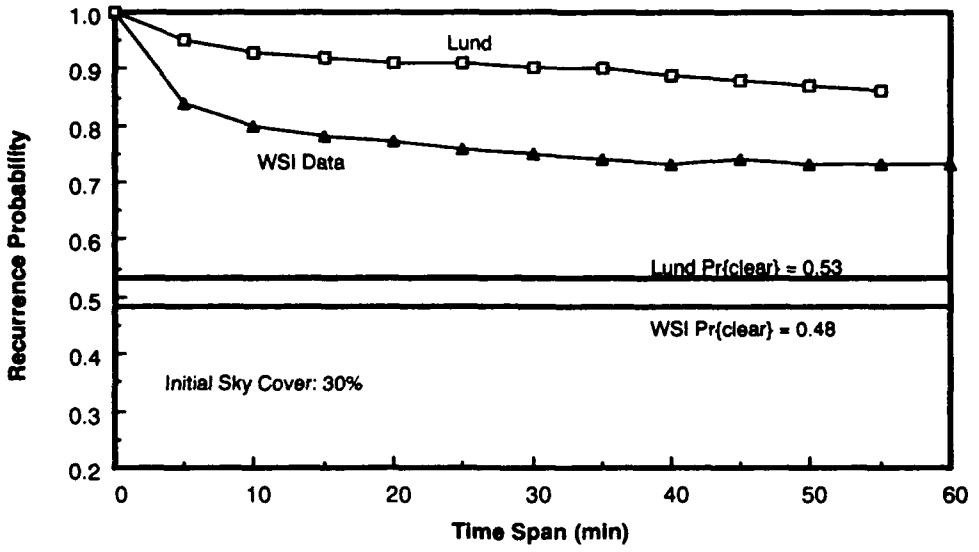


Figure 4.5-3 CFLOS Recurrence Probability — Lund Model and WSI Data (Columbia)

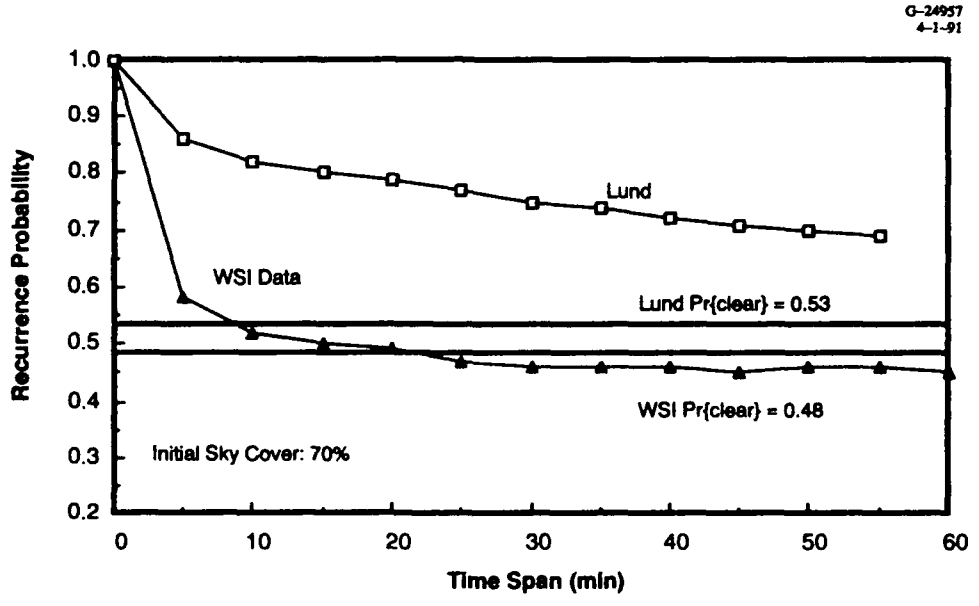


Figure 4.5-4 CFLOS Recurrence Probability — Lund Model and WSI Data (Columbia)

time span increases from 0 to 10 minutes. This is again consistent with the finding that the temporal correlation characterizing the WSI data set is lower than that of the model and the Lund data set.

5. CONCLUSIONS AND RECOMMENDATIONS

Validation of the CFLOS4D simulator, using the limited WSI data set at hand, has been completed. A data requirements analysis for the CFLOS simulator validation effort, which assesses improvements in statistical confidence and power if a larger WSI data set were to become available, has also been completed. This final report has provided an overview of our approach for

- *system-level* validation (based on rigorous statistical comparisons of CFLOS4D and WSI day-to-day downtime count and percentage downtime histograms via the Kolmogorov-Smirnov goodness-of-fit test)
- *model component* investigations (based on multi-faceted examination of various subprograms underlying the CFLOS simulator hierarchy).

Specific conclusions of these analyses include:

- the CFLOS4D simulator underpredicts short duration downtimes (i.e., in the 1-5 minute category) with 95% confidence even if one is willing to tolerate a 30% discrepancy between CFLOS4D and WSI means
- the CFLOS4D simulator overpredicts long duration downtimes (i.e., in the 6-30 minute, 31-180 minute or greater-than-three hour categories); this is especially true for the 6-30 minute category, for which overprediction is evident with 95% confidence even if one is willing to tolerate a 40% discrepancy between CFLOS4D and WSI means
- WSI temporal correlation estimates suggest that model temporal correlation values are too high; adjusting relevant parameters in the model has been demonstrated to yield system-level agreement between model and data.

The conclusion regarding temporal correlation provides an illustration of how fine-tuning a specific model component can improve overall system performance. In this case, the tendency for the CFLOS4D simulator to overpredict long duration downtimes and underpredict short duration downtimes is abated by reducing the strength of temporal dependence.

Given current funding limitations, we recommend that instead of terminating WSI data collection altogether, a reduced processing load be considered. (Extracting data from only, say, 65 selected sky dome points in each image would reduce the deliverable cloud/no-cloud data volume by three orders of magnitude and yet would still allow TASC to carry out key model validation tasks.) TASC further recommends that funding continue so that a validation procedure utilizing multiple pixels per WSI image

can be developed and applied to the existing data set. For a scenario involving several sites, the increase in statistical power is expected to be considerable even with significant inter-pixel sky dome correlations present. Further investigation of model performance given the model parameter adjustment successfully applied in Section 4.4 is also warranted.

REFERENCES

1. Burger, C. F., World Atlas of Total Sky-Cover, Air Force Geophysics Laboratory, AFGL-TR-85-0198, September 1985, ADA 170474.
2. Cloud-Free Line-of-Sight (CFLOS) Simulation Models (User's Manual), USAFETAC/PR-86/002, May 1986.
3. Boehm, A., Gringorten, I., Burger, C., Simulated Duration of CFLOS From Multiple Sites to a Satellite, Air Force Geophysics Laboratory, internal AFGL technical memorandum.
4. Shields, J. E., Koehler, T. L., and Johnson, R. W., Whole Sky Imager, Proceedings of the Cloud Impacts on DoD Operations and Systems — 1989/90 Conference (CIDOS - 89/90), pp. 123-128, January 1990.
5. Finch, S. R., and MacNichol, K. B., Preliminary Data Requirements Analysis for CFLOS4D/CFARC Model Validation, Proceedings of the Cloud Impacts on DoD Operations and Systems — 1989/90 Conference (CIDOS - 89/90), pp. 165-171, January 1990.
6. MacNichol, K. B., Preliminary Data Requirements Analysis for CFLOS4D/CFARC Model Validation: Overview, TASC Technical Information Memorandum TIM-5674-1, March 1989.
7. Finch, S. R., Preliminary Data Requirements Analysis for CFLOS4D/CFARC Model Validation: Detailed Report, TASC Technical Information Memorandum TIM-5674-2, March 1989.
8. Finch, S. R., Data Requirements Analysis Approach for CFLOS4D/CFARC Model Validation, TASC Technical Information Memorandum TIM-5674-3, November 1989.
9. Lund, I. A., Persistence and Recurrence Probabilities of Cloud-Free and Cloudy Lines-of-Sight Through the Atmosphere, J. Appl. Meteor., vol. 12, pp.1222-1228.
10. Medler, C. L., and MacNichol, K. B., Comment on 'Persistence and Recurrence Probabilities of Cloud-Free and Cloudy Lines-of-Sight Through the Atmosphere,' accepted for publication in J. Appl. Meteor.
11. Gleser, L. J., Exact Power of Goodness-of-Fit Tests of Kolmogorov Type For Discontinuous Distributions, J. Amer. Stat. Assoc., vol. 80, pp. 954-958, December 1985.
12. Wood, C. L., and Altavela, M. M., Large-Sample Results for Kolmogorov-Smirnov Statistics for Discrete Distributions, Biometrika, vol. 65, pp. 235-239, January 1978.
13. Allen, J. H., and Malick, J. D., The Frequency of Cloud-Free Viewing Intervals, AIAA Twenty First Aerospace Sciences Meeting, January 1983.
14. Lund, I. A., and Shanklin, M. D., Universal Methods for Estimating Probabilities of Cloud-Free Lines-of-Sight Through the Atmosphere, J. Appl. Meteor., vol. 12, pp. 28-35, February 1973.
15. Anderson, T. W., An Introduction to Multivariate Statistical Analysis, John Wiley and Sons, New York, 1958.